

# Clarke subgradients for directionally Lipschitzian stratifiable functions

D. Drusvyatskiy\*

A. D. Ioffe†

A. S. Lewis‡

November 16, 2012

## Abstract

Using a geometric argument, we show that under a reasonable continuity condition, the Clarke subdifferential of a semi-algebraic (or more generally stratifiable) directionally Lipschitzian function admits a simple form: the normal cone to the domain and limits of gradients generate the entire Clarke subdifferential. The characterization formula we obtain unifies various apparently disparate results that have appeared in the literature. Our techniques also yield a simplified proof that closed semialgebraic functions on  $\mathbf{R}^n$  have a limiting subdifferential graph of uniform local dimension  $n$ .

## 1 Introduction.

Variational analysis, a subject that has been vigorously developing for the past 40 years, has proven itself to be extremely effective at describing nonsmooth phenomenon. The Clarke subdifferential (or generalized gradient) and the limiting subdifferential of a function are the earliest and most widely used constructions of the subject. A key distinction between these two notions is that, in contrast to the limiting subdifferential, the Clarke subdifferential is always convex. From a computational point of view, convexity of the Clarke subdifferential is a great virtue. To illustrate, by the classical Rademacher theorem, a *locally Lipschitz continuous* function  $f$  on an open subset  $U$  of  $\mathbf{R}^n$  is differentiable almost everywhere on  $U$ , in the sense of Lebesgue measure. Clarke, in [5], showed that for such functions, the Clarke subdifferential admits the simple presentation

$$\partial_c f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\}, \quad (\text{LipR})$$

where  $\bar{x}$  is any point of  $U$  and  $\Omega$  is any full measure subset of  $U$ . Such a formula holds great computational promise since gradients are often cheap to compute. For example, utilizing (LipR), Burke, Lewis, and Overton developed an effective computational scheme for approximating the Clarke subdifferential by sampling gradients [3], and, motivated by this idea, developed a robust optimization algorithm [4].

The authors of [3] further extended Clarke's result to the class of *finite-valued, continuous* functions  $f: U \rightarrow \mathbf{R}$ , defined on an open subset  $U$  of  $\mathbf{R}^n$ , that are *absolutely continuous on lines*, and are *directionally Lipschitzian*; the latter means that the Clarke normal cone to the epigraph of  $f$  is pointed. Under these assumptions on  $f$ , the authors derived the representation

$$\partial_c f(\bar{x}) = \bigcap_{\delta > 0} \text{cl conv} \left( \nabla f(\Omega \cap B_\delta(\bar{x})) \right), \quad (\text{ACLR})$$

---

\*School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; <http://people.orie.cornell.edu/dd379/>. Work of Dmitriy Drusvyatskiy on this paper has been partially supported by the NDSEG grant from the Department of Defense and by the BSF Travel Grant for Young Scientists.

†Department of Mathematics, Technion-Israel Institute of Technology, Haifa, Israel 32000; <http://www.math.technion.ac.il/Site/people/process.php?id=672>. Work of A. D. Ioffe was supported in part by the US-Israel Binational Science Foundation Grant 2008261.

‡School of Operations Research and Information Engineering, Cornell University, Ithaca, New York, USA; <http://people.orie.cornell.edu/aslewis/>. Research supported in part by National Science Foundation Grant DMS-0806057 and by the US-Israel Binational Scientific Foundation Grant 2008261.

where  $B_\delta(\bar{x})$  is an open ball of radius  $\delta$  around  $\bar{x}$  and  $\Omega$  is any full measure subset of  $U$ , and they extended their computational scheme to this more general setting. One can easily see that this formula generalizes Clarke’s result, since locally Lipschitz functions are absolutely continuous on lines, and for such functions (ACLR) reduces to (LipR). Pointedness of the Clarke normal cone is a common theoretical assumption. For instance, closed convex sets with nonempty interior have this property. Some results related to (ACLR) appear in [13].

In optimization theory, one is often interested in *extended real-valued* functions (functions that are allowed to take on the value  $+\infty$ ), so as to model constraints, for instance. Results above are not applicable in such instances. An early predecessor of (LipR) and (ACLR) does rectify this problem, at least when convexity is present. Rockafellar [27, Theorem 25.6] showed that for any closed *convex* function  $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ , whose domain  $\text{dom } f$  has a nonempty interior, the convex subdifferential has the form

$$\partial f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \rightarrow \bar{x} \right\} + N_{\text{dom } f}(\bar{x}), \quad (\text{CoR})$$

where  $\bar{x}$  is any point in the domain of  $f$  and  $N_{\text{dom } f}(\bar{x})$  is the normal cone to the domain of  $f$  at  $\bar{x}$ .

Our goal is to provide an intuitive and geometric proof of a representation formula unifying (LipR), (ACLR), and (CoR). To do so, we will impose a certain structural assumption on the functions  $f$  that we consider. Namely, we will assume that the domain of  $f$  can be locally “stratified” into a finite collection of smooth manifolds, so that  $f$  is smooth on each such manifold. Many functions of practical importance in optimization and in nonsmooth analysis possess this property. All semi-algebraic functions (those functions whose graphs can be described as a union of finitely many sets, each defined by finitely many polynomial inequalities), and more generally, tame functions fall within this class [18]. We will show (Theorem 3.10) that for a *directionally Lipschitzian, stratifiable* function  $f: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$ , that is continuous on its domain (for simplicity), the Clarke subdifferential admits the intuitive form

$$\partial_c f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\} + \text{coco} \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\} + N_{\text{dom } f}^c(\bar{x}), \quad (1)$$

or equivalently,

$$\partial_c f(\bar{x}) = \bigcap_{\delta > 0} \text{cl conv} \left( \nabla f(\Omega \cap B_\delta(\bar{x})) \right) + N_{\text{dom } f}^c(\bar{x}),$$

where  $\Omega$  is any dense subset of  $\text{dom } f$  and  $\text{coco}$  denotes the convex conical hull. (In contrast to the aforementioned results, we do not require  $\Omega$  to have full-measure).

This is significant both from theoretical and computational perspectives. Proofs of (LipR) and (ACLR) are based largely on Fubini’s theorem and analysis of directional derivatives, and though the arguments are elegant, they do not shed light on the geometry driving such representations to hold. Similarly, Rockafellar’s argument of (CoR) relies heavily on the well-oiled machinery of convex analysis. Consequently, a simple unified geometric argument is extremely desirable. From a practical point of view, representation (1) decouples the behavior of the function from the geometry of the domain; consequently, when the domain is a simple set (polyhedral perhaps) and the behavior of the function on the interior of the domain is complex, our result provides a convenient method of calculating the Clarke subdifferential purely in terms of limits of gradients and the normal cone to the domain — information that is often readily available. Furthermore, using (1), the functions we consider in the current paper become amenable to the techniques developed in [3].

Whereas (1) deals with pointwise estimation of the Clarke subdifferential, our second result addresses the geometry of subdifferential graphs, as a whole. In particular, we consider the size of subdifferential graphs, a feature that may have important algorithmic applications. For instance, Robinson [26, 25] shows computational promise for functions defined on  $\mathbf{R}^n$  whose subdifferential graphs are locally homeomorphic to an open subset of  $\mathbf{R}^n$ . Due to the results of Minty [20] and Poliquin-Rockafellar [24], Robinson’s techniques are applicable for convex, and more generally, for “prox-regular” functions. Trying to understand the size of subdifferential graphs in the absence of convexity (or monotonicity), the authors in [10] were led to consider the semi-algebraic setting. The authors proved that the *limiting subdifferential graph* of a closed, proper, *semi-algebraic* function on  $\mathbf{R}^n$  has *uniform local dimension*  $n$ . Applications to sensitivity analysis were also discussed. We show how the techniques developed in the current paper

drastically simplify the proof of this striking fact. Remarkably, this dimensional uniformity does not hold for the Clarke subdifferential graph.

The rest of the paper is organized as follows. In Section 2, we establish notation and recall some basic facts from variational analysis. In Section 3, we derive a characterization formula for the Clarke subdifferential of a directionally Lipschitzian, stratifiable function that possesses a certain continuity property on its domain, and in Section 4 we relate our results to the gradient sampling framework. In Section 5, we prove the theorem concerning the local dimension of semi-algebraic subdifferential graphs. We have designed this last section to be entirely independent from the previous ones (except for Section 2), since it does require a short foray into semi-algebraic geometry.

## 2 Preliminary results.

In this section, we summarize some of the fundamental tools used in variational analysis and nonsmooth optimization. We refer the reader to the monographs Borwein-Zhu [2], Clarke-Ledyaev-Stern-Wolenski [6], Mordukhovich [21, 22], and Rockafellar-Wets [29], for more details. Unless otherwise stated, we follow the terminology and notation of [29].

The functions we consider will take their values in the extended real line  $\overline{\mathbf{R}} := \mathbf{R} \cup \{-\infty, \infty\}$ . We say that an extended-real-valued function is *proper* if it is never  $-\infty$  and is not always  $+\infty$ . For a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , the *domain* of  $f$  is

$$\text{dom } f := \{x \in \mathbf{R}^n : f(x) < +\infty\},$$

and the *epigraph* of  $f$  is

$$\text{epi } f := \{(x, r) \in \mathbf{R}^n \times \mathbf{R} : r \geq f(x)\}.$$

Throughout this work, we will only use Euclidean norms. Hence for a point  $x \in \mathbf{R}^n$ , the symbol  $|x|$  will denote the standard Euclidean norm of  $x$ . Unless we state otherwise, the topology on  $\mathbf{R}^n$  that we consider is induced by this norm. Given a set  $Q \subset \mathbf{R}^n$ , the notation  $x_i \xrightarrow{Q} \bar{x}$  will mean that the sequence  $x_i$  converges to  $\bar{x}$  and all the points  $x_i$  lie in  $Q$ . We let “ $o(|x - \bar{x}|)$  for  $x \in Q$ ” be shorthand for a function that satisfies  $\frac{o(|x - \bar{x}|)}{|x - \bar{x}|} \rightarrow 0$  whenever  $x \xrightarrow{Q} \bar{x}$  with  $x \neq \bar{x}$ .

A function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is *locally Lipschitz continuous* at a point  $\bar{x}$  relative to a set  $Q \subset \mathbf{R}^n$  containing  $\bar{x}$ , if  $f(\bar{x})$  is finite and there exists a real number  $\kappa \in [0, +\infty)$  with

$$|f(x) - f(y)| \leq \kappa|x - y|, \text{ for all } x, y \in Q \text{ near } \bar{x}.$$

If there exists an open neighborhood  $Q$  of  $\bar{x}$  so that the above conditions hold, then we simply say that  $f$  is locally Lipschitz continuous at  $\bar{x}$ .

A *set-valued mapping*  $F$  from  $\mathbf{R}^n$  to  $\mathbf{R}^m$ , denoted by  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ , is a mapping from  $\mathbf{R}^n$  to the power set of  $\mathbf{R}^m$ . Hence for each point  $x \in \mathbf{R}^n$ ,  $F(x)$  is a subset of  $\mathbf{R}^m$ . For a set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ , the *domain* of  $F$  is

$$\text{dom } F := \{x \in \mathbf{R}^n : F(x) \neq \emptyset\},$$

and the *graph* of  $F$  is

$$\text{gph } F := \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^m : y \in F(x)\}.$$

The *outer limit* of  $F$  at  $\bar{x}$  is

$$\text{Limsup}_{x \rightarrow \bar{x}} F(x) := \{v \in \mathbf{R}^m : \exists x_i \rightarrow \bar{x}, \exists v_i \rightarrow v \text{ with } v_i \in F(x_i)\}.$$

The mapping  $F$  is *locally bounded* near  $\bar{x}$  if the image  $F(V) \subset \mathbf{R}^m$  is bounded, for some neighborhood  $V$  of  $\bar{x}$ . The following definition extends in two ways the classical notion of continuity to set-valued mappings.

**Definition 2.1** (Continuity). Consider a set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ .

1.  $F$  is *outer semicontinuous* at a point  $\bar{x} \in \mathbf{R}^n$  if for any sequence of points  $x_i \in \mathbf{R}^n$  converging to  $\bar{x}$  and any sequence of vectors  $v_i \in F(x_i)$  converging to  $\bar{v}$ , we must have  $\bar{v} \in F(\bar{x})$ .
2.  $F$  is *inner semicontinuous* at  $\bar{x}$  if for any sequence of points  $x_i$  converging to  $\bar{x}$  and any vector  $\bar{v} \in F(\bar{x})$ , there exist vectors  $v_i \in F(x_i)$  converging to  $\bar{v}$ .

If both properties hold, then we say that  $F$  is *continuous* at  $\bar{x}$ .

We let  $\text{int } Q$ ,  $\text{cl } Q$ ,  $\text{conv } Q$ , and  $\text{coco } Q$ , denote the interior, closure, convex hull, and convex conical hull of a set  $Q$ , respectively. A cone  $Q$  is said to be *pointed* if it contains no lines. An open ball of radius  $r$  around a point  $\bar{x} \in \mathbf{R}^n$  will be denoted by  $B_r(\bar{x})$ . We let  $\mathbf{B}$  and  $\bar{\mathbf{B}}$  be the open and closed unit balls, respectively. The following is a standard result on preservation of continuity of set-valued mappings under a pointwise convex conical hull operation. We provide a proof for completeness.

**Lemma 2.2** (Preservation of Continuity). *Consider a set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  that is outer-semicontinuous at a point  $\bar{x} \in \mathbf{R}^n$ . Suppose  $F$  is locally bounded near  $\bar{x}$  and  $0 \notin F(\bar{x})$ . Then the mapping*

$$x \mapsto G(x) := \text{coco } F(x)$$

*is outer-semicontinuous at  $\bar{x}$ , provided that  $G(\bar{x})$  is pointed.*

*Proof.* Consider a sequence  $(x_i, v_i) \rightarrow (\bar{x}, \bar{v})$ , with  $v_i \in G(x_i)$  for each index  $i$ . By Carathéodory's theorem, we deduce

$$v_i = \sum_{j=1}^m \lambda_j^i y_j^i,$$

for some multipliers  $\lambda_j^i \geq 0$  and vectors  $y_j^i \in F(x_i)$ , where  $j = 1, \dots, m$ .

Restricting to a subsequence, we may assume that there exist nonzero vectors  $y_j \in F(\bar{x})$  satisfying

$$y_j = \lim_{i \rightarrow \infty} y_j^i, \text{ for each index } j.$$

We claim that the sequence of multipliers  $\lambda_j^i$  is bounded. Indeed suppose this is not the case and let  $m_i := \max_j \lambda_j^i$ . Then up to a subsequence, there exist multipliers  $\lambda_j$ , not all zero, such that

$$\frac{\lambda_j^i}{m_i} \rightarrow \lambda_j \text{ as } i \rightarrow \infty, \text{ and } 0 = \sum_{j=1}^m \lambda_j y_j,$$

contradicting the fact that  $G(\bar{x})$  is pointed. We conclude that the multipliers  $\lambda_j^i$  are bounded.

Then up to a subsequence, we have

$$\bar{v} = \lim_{i \rightarrow \infty} \sum_{j=1}^m \lambda_j^i y_j^i = \sum_{j=1}^m \lambda_j y_j \in G(\bar{x}),$$

for some real numbers  $\lambda_j \geq 0$ . We conclude that  $G$  is outer-semicontinuous at  $\bar{x}$ . □

The distance of a point  $x$  to a set  $Q$  is

$$d_Q(x) := \inf_{y \in Q} |x - y|,$$

and the projection of  $x$  onto  $Q$  is

$$P_Q(x) := \{y \in Q : |x - y| = d_Q(x)\}.$$

We now consider normal cones, which are fundamental objects in variational geometry.

**Definition 2.3** (Proximal normals). Consider a set  $Q \subset \mathbf{R}^n$  and a point  $\bar{x} \in Q$ . The *proximal normal cone* to  $Q$  at  $\bar{x}$ , denoted  $N_Q^P(\bar{x})$ , consists of all vectors  $v \in \mathbf{R}^n$  such that  $\bar{x} \in P_Q(\bar{x} + \frac{1}{r}v)$  for some  $r > 0$ .

Geometrically, a vector  $v \neq 0$  is a proximal normal to  $Q$  at  $\bar{x}$  precisely when there exists a ball touching  $Q$  at  $\bar{x}$  such that  $v$  points from  $\bar{x}$  towards the center of the ball. Furthermore, this condition amounts to

$$\langle v, x - \bar{x} \rangle \leq O(|x - \bar{x}|^2) \quad \text{as } x \rightarrow \bar{x} \text{ in } Q.$$

Relaxing the inequality above, one obtains the following notion.

**Definition 2.4** (Frechét normals). Consider a set  $Q \subset \mathbf{R}^n$  and a point  $\bar{x} \in Q$ . The *Frechét normal cone* to  $Q$  at  $\bar{x}$ , denoted  $\hat{N}_Q(\bar{x})$ , consists of all vectors  $v \in \mathbf{R}^n$  such that

$$\langle v, x - \bar{x} \rangle \leq o(|x - \bar{x}|) \quad \text{as } x \rightarrow \bar{x} \text{ in } Q.$$

Note that both  $N_Q^P(\bar{x})$  and  $\hat{N}_Q(\bar{x})$  are convex cones, while  $\hat{N}_Q(\bar{x})$  is also closed. Evidently, the set-valued mapping  $x \mapsto \hat{N}_Q(x)$  is generally not outer-semicontinuous, and hence is not robust relative to perturbations in  $x$ . To correct for that, the following definition is introduced.

**Definition 2.5** (Limiting normals). Consider a set  $Q \subset \mathbf{R}^n$  and a point  $\bar{x} \in Q$ . The *limiting normal cone* to  $Q$  at  $\bar{x}$ , denoted  $N_Q(\bar{x})$ , consists of all vectors  $v \in \mathbf{R}^n$  such that there are sequences  $x_i \xrightarrow{Q} \bar{x}$  and  $v_i \rightarrow v$  with  $v_i \in \hat{N}_Q(x_i)$ .

The limiting normal cone, as defined above, consists of limits of Frechét normals. In fact, the same object arises if we only take limits of proximal normals [29, Exercise 6.18]. Convexifying the limiting normal cone leads to the following definition.

**Definition 2.6** (Clarke normals). Consider a set  $Q \subset \mathbf{R}^n$  and a point  $\bar{x} \in Q$ . The *Clarke normal cone* to  $Q$  at  $\bar{x}$  is

$$N_Q^c(\bar{x}) := \text{cl conv } N_Q(\bar{x}).$$

Given any set  $Q \subset \mathbf{R}^n$  and a mapping  $F: Q \rightarrow \tilde{Q}$ , where  $\tilde{Q} \subset \mathbf{R}^m$ , we say that  $F$  is  $\mathbf{C}^p$ -smooth ( $p \geq 2$ ) if for each point  $\bar{x} \in Q$ , there is a neighborhood  $U$  of  $\bar{x}$  and a  $\mathbf{C}^p$  mapping  $\hat{F}: \mathbf{R}^n \rightarrow \mathbf{R}^m$  that agrees with  $F$  on  $Q \cap U$ . Henceforth, to simplify notation, the word smooth will mean  $\mathbf{C}^2$ -smooth.

**Definition 2.7** (Smooth Manifolds). We say that a set  $M$  in  $\mathbf{R}^n$  is a  $\mathbf{C}^2$ -submanifold of dimension  $r$  if for each point  $\bar{x} \in M$ , there is an open neighborhood  $U$  around  $\bar{x}$  and a mapping  $F: \mathbf{R}^n \rightarrow \mathbf{R}^{n-r}$  that is  $\mathbf{C}^2$ -smooth with  $\nabla F(\bar{x})$  of full rank, satisfying  $M \cap U = \{x \in U : F(x) = 0\}$ .

A good reference on smooth manifold theory is [19].

**Theorem 2.8.** [29, Example 6.8] Consider a  $\mathbf{C}^2$ -manifold  $M \subset \mathbf{R}^n$ . Then at every point  $x \in M$ , the normal cone  $N_M(x)$  is equal to the normal space to  $M$  at  $x$ , in the sense of differential geometry.

In fact, the following stronger characterization holds.

**Theorem 2.9** (Prox-normal neighborhood). Consider a  $\mathbf{C}^2$ -manifold  $M \subset \mathbf{R}^n$  and a point  $\bar{x} \in M$ . Then there exists an open neighborhood  $U$  of  $\bar{x}$ , such that

1. the projection map  $P_M$  is single-valued on  $U$ ,
2. for any two points  $x \in M \cap U$  and  $v \in U$ , the equivalence,

$$v \in x + N_M(x) \Leftrightarrow x = P_M(v),$$

holds.

Following the notation of [14], we call the set  $U$  that is guaranteed to exist by Theorem 2.9, a *prox-normal neighborhood* of  $M$  at  $\bar{x}$ . For more details about Theorem 2.9, see [29, Exercise 13.38], [6, Proposition 1.9]. We should note that the theorem above holds for all “prox-regular” sets  $M$  [24].

Armed with the aforementioned facts from variational geometry, we can study variational properties of functions via their subdifferential mappings.

**Definition 2.10** (Subdifferentials). Consider a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and a point  $\bar{x} \in \mathbf{R}^n$ , with  $f(\bar{x})$  finite. The *limiting subdifferential* of  $f$  at  $\bar{x}$  is defined by

$$\partial f(\bar{x}) = \{v \in \mathbf{R}^n : (v, -1) \in N_{\text{epi } f}(\bar{x}, f(\bar{x}))\}.$$

*Proximal*, *Frechét*, and *Clarke subdifferentials* are defined analogously.

For  $\bar{x}$  such that  $f(\bar{x})$  is not finite, we follow the convention that  $\partial_P f(\bar{x}) = \hat{\partial} f(\bar{x}) = \partial f(\bar{x}) = \partial_c f(\bar{x}) = \emptyset$ .

The subdifferentials defined above fail to capture the horizontal normals to the epigraph. Hence to obtain a more complete picture, we consider the following.

**Definition 2.11** (Horizon subdifferential). For a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is finite at a point  $\bar{x}$ , the *horizon subdifferential* is given by

$$\partial^\infty f(\bar{x}) = \{v \in \mathbf{R}^n : (v, 0) \in N_{\text{epi } f}(\bar{x}, f(\bar{x}))\}.$$

For a set  $Q \subset \mathbf{R}^n$ , we define  $\delta_Q: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  to be a function that is 0 on  $Q$  and  $+\infty$  elsewhere. We call  $\delta_Q$  the *indicator function* of  $Q$ . Then for a point  $\bar{x}$ , we have  $N_Q(\bar{x}) = \partial \delta_Q(\bar{x})$ , with analogous statements holding for the other subdifferentials.

Often, we will work with discontinuous functions  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . For such functions, it is useful to consider  $f$ -*attentive* convergence of a sequence  $x_i$  to a point  $\bar{x}$ , denoted  $x_i \xrightarrow{f} \bar{x}$ . In this notation we have

$$x_i \xrightarrow{f} \bar{x} \iff x_i \rightarrow \bar{x} \text{ and } f(x_i) \rightarrow f(\bar{x}).$$

If in addition we have a set  $Q \subset \mathbf{R}^n$ , then  $x_i \xrightarrow{Q} \bar{x}$  will mean that  $x_i$  converges  $f$ -attentively to  $\bar{x}$  and the points  $x_i$  all lie in  $Q$ . It is immediate that the mappings  $\partial f$  and  $\partial^\infty f$  are outer-semicontinuous with respect to  $f$ -attentive convergence  $x \xrightarrow{f} \bar{x}$ .

Consider a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is locally lower semi-continuous at a point  $\bar{x}$ , with  $f(\bar{x})$  finite. Then  $f$  is locally Lipschitz continuous around  $\bar{x}$  if and only if the horizon subdifferential is trivial, that is the condition  $\partial^\infty f(\bar{x}) = \{0\}$  holds [29, Theorem 9.13]. Weakening the latter condition to requiring  $\partial^\infty f(\bar{x})$  to simply be pointed, we arrive at the following central notion [29, Exercise 9.42].

**Definition 2.12** (epi-Lipschitzian sets and directionally Lipschitzian functions).

1. A set  $Q \subset \mathbf{R}^n$  is *epi-Lipschitzian* at one of its points  $\bar{x}$  if  $Q$  is locally closed at  $\bar{x}$  and the normal cone  $N_Q(\bar{x})$  is pointed.
2. A function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ , is *directionally Lipschitzian* at  $\bar{x}$  if  $f$  is locally lower-semicontinuous at  $\bar{x}$  and the cone  $\partial^\infty f(\bar{x})$  is pointed.

Rockafellar [28, Section 4] proved that an epi-Lipschitzian set in  $\mathbf{R}^n$ , up to a rotation, locally coincides with an epigraph of a Lipschitz continuous function defined on  $\mathbf{R}^{n-1}$ . We should further note that the Clarke normal cone mapping of an epi-Lipschitzian set is outer-semicontinuous [6, Proposition 6.8].

It is easy to see that a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is directionally Lipschitzian at  $\bar{x}$  if and only if the epigraph  $\text{epi } f$  is epi-Lipschitzian at  $(\bar{x}, f(\bar{x}))$ . Furthermore, for a set  $Q$  that is locally closed at  $\bar{x}$ , the limiting normal cone  $N_Q(\bar{x})$  is pointed if and only if the Clarke normal cone  $N_Q^c(\bar{x})$  is pointed [29, Exercise 9.42].

Consider the two functions

$$f_1(x) = x \quad \text{and} \quad f_2(x) = \begin{cases} x & \text{if } x \leq 0 \\ x + 1 & \text{if } x > 0 \end{cases}$$

defined on the real line. Clearly both  $f_1$  and  $f_2$  are directionally Lipschitzian, and have the same derivatives at each point of differentiability. However  $\partial_c f_1(0) \neq \partial_c f_2(0)$ . Roughly speaking, this situation arises because some normal cones to the epigraph of a function  $f$ , namely at points  $(x, r)$  with  $r > f(x)$ , may not correspond to any subdifferential. Consequently, if we have any hope of deriving a characterization of the Clarke subdifferential purely in terms of gradients and the normal cone to the domain, we must

eliminate the situation above. Evidently, assumption of continuity of the function on the domain would do the trick. However, such an assumption would immediately eliminate many interesting convex functions from consideration. Rather than doing so, we identify a new condition, which arises naturally as a byproduct of our arguments. At the risk of sounding extravagant, we give this property a name.

**Definition 2.13** (Vertical continuity). We say that a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is *vertically continuous* at a point  $\bar{x} \in \text{dom } f$  if the equation

$$\limsup_{\substack{x_i \rightarrow \bar{x}, \, r \rightarrow f(\bar{x}) \\ r > f(\bar{x})}} N_{\text{epi } f}(x, r) = N_{\text{dom } f}(\bar{x}) \times \{0\}, \quad (2)$$

holds.

To put this condition in perspective, we record the following observations.

**Proposition 2.14** (Properties of vertically continuous functions). *Consider a proper function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is locally lower-semicontinuous at a point  $\bar{x}$ , with  $f(\bar{x})$  finite.*

1. *Suppose that whenever a pair  $(x, r) \in \text{epi } f$ , with  $r > f(x)$ , is near  $(\bar{x}, f(\bar{x}))$  we have*

$$N_{\text{epi } f}(x, r) = N_{\text{dom } f}(x) \times \{0\}.$$

*Then  $f$  is vertically continuous at  $\bar{x}$ .*

2. *Suppose that  $\bar{x}$  lies in the interior of  $\text{dom } f$  and that  $f$  is vertically continuous at  $\bar{x}$ . Then  $f$  is continuous at  $\bar{x}$ , in the usual sense.*
3. *Suppose that  $f$  is continuous on a neighborhood of  $\bar{x}$ , relative to the domain of  $f$ . Then  $f$  is vertically continuous at all points of  $\text{dom } f$  near  $\bar{x}$ .*
4. *If  $f$  is convex, then  $f$  is vertically continuous at every point  $\bar{x}$  in  $\text{dom } f$ .*
5. *Suppose that  $f$  is “amenable” at  $\bar{x}$  in the sense of [23]; that is,  $f$  is finite at  $\bar{x}$  and there exists a neighborhood  $V$  of  $\bar{x}$  so that  $f$  can be written as a composition  $f = g \circ F$ , for a  $\mathbf{C}^1$  mapping  $F: V \rightarrow \mathbf{R}^m$  and a proper, lower-semicontinuous, convex function  $g: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$ , so that the qualification condition*

$$N_{\text{dom } g}(F(\bar{x})) \cap \ker \nabla F(\bar{x})^* = \{0\}, \quad (3)$$

*is satisfied. Then  $f$  is vertically continuous at  $\bar{x}$ .*

*Proof.* Claim 1 is immediate from outer-semicontinuity of the normal cone map  $N_{\text{dom } f}$ .

To see 2, suppose that  $f$  is vertically continuous at  $\bar{x} \in \text{int dom } f$ . Since the normal cone to the domain of  $f$  at  $\bar{x}$  consists of the zero vector, we deduce  $N_{\text{epi } f}(\bar{x}, f(\bar{x}) + \frac{1}{n}) = \{0\}$ , for all sufficiently large integers  $n$ . By [29, Exercise 6.19], we deduce that each such point  $(\bar{x}, f(\bar{x}) + \frac{1}{n})$  lies in the interior of the epigraph  $\text{epi } f$ . Consequently for any sequence  $x_i \rightarrow \bar{x}$ , we have

$$\limsup_{i \rightarrow \infty} f(x_i) \leq f(\bar{x}) + \frac{1}{n},$$

for all large indices  $n$ . Letting  $n$  tend to infinity, we deduce that  $f$  is upper-semicontinuous at  $\bar{x}$ . The result follows.

To see 3, suppose that  $f$  is continuous at  $x \in \text{dom } f$ , relative to the domain of  $f$ . Then for any real  $r > f(x)$  there exists an  $\epsilon > 0$  so that the epigraph  $\text{epi } f$  coincides with the product set,  $\text{dom } f \times [r - \epsilon, r + \epsilon]$ , locally around  $(x, r)$ . In fact, this follows just from upper-semicontinuity of  $f$  at  $x$ , relative to  $\text{dom } f$ . We deduce  $N_{\text{epi } f}(x, r) = N_{\text{dom } f}(x) \times \{0\}$ . The result follows by 1.

To see 4, consider a pair  $(\bar{x}, r)$  with  $r > f(\bar{x})$  and observe

$$\begin{aligned} (v, \alpha) \in N_{\text{epi } f}(\bar{x}, r) &\iff \langle (v, \alpha), (\bar{x}, r) \rangle \geq \langle (v, \alpha), (x', r') \rangle \quad \text{for all } (x', r') \in \text{epi } f \\ &\iff \alpha = 0 \text{ and } v \in N_{\text{dom } f}(\bar{x}). \end{aligned}$$

Appealing to 1, we obtain the result.

The proof of 5 follows from a standard nonsmooth chain rule. We outline the argument below. Without loss of generality, we can assume that the representation  $f = g \circ F$  holds on all of  $\mathbf{R}^n$ . Observe

$$\text{epi } f = \{(x, r) \in \mathbf{R}^{n+1} : G(x, r) \in \text{epi } g\},$$

for the mapping  $G: \mathbf{R}^{n+1} \rightarrow \mathbf{R}^{m+1}$  defined by  $G(x, r) = (F(x), r)$ . We would like to use the chain rule appearing in [29, Theorem 6.14] to compute the normal cone to  $\text{epi } f$ . To this end, consider a pair  $(x, r) \in \text{epi } f$  and a vector  $(y, \alpha) \in N_{\text{epi } g}(G(x, r))$ . We have

$$\begin{aligned} 0 = \nabla G(x, r)^*(y, \alpha) &\iff \alpha = 0 \text{ and } \nabla F(x)^*y = 0 \\ &\iff y \in N_{\text{dom } g}(F(x)) \text{ and } \nabla F(x)^*y = 0 \\ &\iff y = 0, \end{aligned}$$

where the last equivalence follows from the qualification condition (3). Applying the chain rule [29, Theorem 6.14], we deduce

$$N_{\text{epi } f}(x, r) = \nabla G(x, r)^* N_{\text{epi } g}(G(x, r)),$$

for all pairs  $(x, r) \in \text{epi } f$ . In particular, if we have  $r > f(x)$ , or equivalently  $r > g(F(x))$ , we deduce

$$N_{\text{epi } f}(x, r) = \nabla F(x)^* N_{\text{dom } g}(F(x)) \times \{0\}.$$

The right hand side coincides with  $N_{\text{dom } f}(x) \times \{0\}$  by [23, Theorem 3.3]. The result follows by appealing to 1.  $\square$

As can be seen from the proposition above, vertical continuity bridges the gap between continuity of the function on the interior of the domain and continuity on the whole domain, and hence the name. In summary, all convex and amenable functions have this property, as do functions that are continuous on their domains. An illustrative example is provided by the proper, lower semi-continuous, convex (directionally Lipschitzian) function  $f$  on  $\mathbf{R}^2$ , defined by

$$f(x, y) = \begin{cases} y^2/2x & \text{if } x > 0 \\ 0 & \text{if } x = 0, y = 0 \\ \infty & \text{otherwise} \end{cases}$$

This function is discontinuous at the origin, despite being vertically continuous there.

In the sequel, we will need the following basic result.

**Proposition 2.15.** *Consider a set  $M \subset \mathbf{R}^n$  and a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is finite-valued and smooth on  $M$ . Then, at any point  $\bar{x} \in M$ , we have*

$$\hat{\partial}f(\bar{x}) \subset \nabla g(\bar{x}) + N_M(\bar{x}),$$

where  $g: \mathbf{R}^n \rightarrow \mathbf{R}$  is any smooth function agreeing with  $f$  on  $M$  on a neighborhood of  $\bar{x}$ .

*Proof.* Define a function  $h: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  agreeing with  $f$  on  $M$  and equalling plus infinity elsewhere. It is standard to check that the chain of inclusions,

$$\hat{\partial}f(\bar{x}) \subset \hat{\partial}h(\bar{x}) = \hat{\partial}(g(\cdot) + \delta_M(\cdot))(\bar{x}) \subset \nabla g(\bar{x}) + N_M(\bar{x}),$$

holds.  $\square$

### 3 Characterization of the Clarke Subdifferential.

Directionally Lipschitzian functions play an important role in optimization and are close relatives of locally Lipschitz functions [3, 29]. Indeed, Rockafellar showed that epi-Lipschitzian sets, up to a change of coordinates, are epigraphs of Lipschitz functions [28, Section 4].



As was mentioned in the introduction, a key feature of Clarke's construction is that the Clarke subdifferential of a locally Lipschitz function  $f$ , on  $\mathbf{R}^n$ , can be described purely in terms of gradient information. It is then reasonable to hope that the same property holds for continuous directionally Lipschitzian functions, but this is too good to be true. Though such functions are differentiable almost everywhere [1], their gradients may fail to generate the entire Clarke subdifferential. A simple example is furnished by the classical ternary Cantor function — a nondecreasing, continuous, and therefore directionally Lipschitzian function, with zero derivative at each point of differentiability. The Clarke subdifferential of this function does not identically consist of the zero vector [2, Exercise 3.5.5], and consequently cannot be recovered from classical derivatives. This example notwithstanding, one does not expect the Cantor function to arise often in practice.

Nonsmoothness arises naturally in many applications, but not pathologically so. On the contrary, nonsmoothness is usually highly structured. Often such structure manifests itself through existence of a *stratification*. In the current work, we consider so-called *stratifiable* functions. Roughly speaking, domains of such function can be decomposed into smooth manifolds (called strata), which fit together in a “regular” way, and so that the function is smooth on each such stratum. In particular, this rich class of functions includes all semi-algebraic, and more generally, all o-minimally defined functions. See for example [30]. We now make this notion precise.

**Definition 3.1** (Locally finite stratifications). Consider a set  $Q$  in  $\mathbf{R}^n$ . A *locally finite stratification* of  $Q$  is a partition of  $Q$  into disjoint manifolds  $M_i$  (called strata) satisfying

- **(frontier condition)** for each index  $i$ , the closure of  $M_i$  in  $Q$  is the union of some  $M_j$ 's, and
- **(local finiteness)** each point  $x \in Q$  has a neighborhood that intersects only finitely many strata.

We say that a set  $Q \subset \mathbf{R}^n$  is *stratifiable* if it admits a locally finite stratification.

Observe that due to the frontier condition, a stratum  $M_i$  intersects the closure of another stratum  $M_j$  if and only if the inclusion  $M_i \subset \text{cl } M_j$  holds. Consequently, given a locally finite stratification of a set  $Q$  into manifolds  $\{M_i\}$ , we can impose a natural partial order on the strata, namely

$$M_i \preceq M_j \Leftrightarrow M_i \subset \text{cl } M_j.$$

A good example to keep in mind is the partition of a convex polyhedron into its open faces.

**Definition 3.2** (Stratifiable functions). A function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  is *stratifiable* if there exists a locally finite stratification of  $\text{dom } f$  so that  $f$  is smooth on each stratum.

The following result nicely illustrates the geometric insight one obtains by working with stratifications explicitly.

**Proposition 3.3** (Dense differentiability). Consider a proper stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is directionally Lipschitzian at all points of  $\text{dom } f$  near  $\bar{x}$ , and let  $\Omega$  be any dense subset of  $\text{dom } f$ . Then the set  $\Omega \cap \text{dom } \nabla f$  is dense in the domain of  $f$ , in the  $f$ -attentive sense, locally near  $\bar{x}$ .

*Proof.* Consider a locally finite stratification of  $\text{dom } f$  into manifolds  $M_i$  so that  $f$  is smooth on each stratum. Suppose for the sake of contradiction that there exists a point  $x \in \text{dom } f$  arbitrarily close to  $\bar{x}$  and an  $f$ -attentive neighborhood  $V = \{y \in \mathbf{R}^n : |y - x| < \epsilon, |f(y) - f(x)| < \delta\}$  so that  $V \cap \Omega$  does not intersect any strata of dimension  $n$ . Shrinking  $V$ , we may assume that  $V$  intersects only finitely many strata, say  $\{M_j\}$  for  $j \in J := \{1, \dots, k\}$ , and that the inclusion  $x \in \text{cl } M_j$  holds for each index  $j \in J$ . Notice that since  $f$  is continuous on each stratum, the set  $V$  is a union of open subsets of the strata  $M_j$  for  $j \in J$ .

Now among the strata  $M_j$  with  $j \in J$ , choose a stratum  $M$  that is maximal with respect to the partial order  $\preceq$ . Clearly, we have

$$M \cap \text{cl } M_j = \emptyset, \text{ for each } j \in J \text{ with } M_j \neq M.$$

Now let  $y$  be any point of  $V \cap M$  and observe that there exists a neighborhood  $Y$  of  $y$  so that the functions  $f$  and  $f + \delta_M$  coincide on  $Y \cap M$ . We deduce that  $\partial f(y)$  is a nontrivial affine subspace. Since

$f$  is directionally Lipschitzian at all points in  $\text{dom } f$  near  $\bar{x}$ , and in particular at  $y$ , we have arrived at a contradiction. Thus  $\Omega \cap \text{dom } \nabla f$  is dense (in the  $f$ -attentive sense) in the domain of  $f$ , locally near  $\bar{x}$ .  $\square$

In this section, we will derive a characterization formula for the Clarke subdifferential of a stratifiable, vertically continuous, directionally Lipschitzian function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . This formula will only depend on the gradients of  $f$  and on the normal cone to the domain. It is important to note that the characterization formula, we obtain, is independent of any particular stratification of  $\text{dom } f$ ; one only needs to know that  $f$  is stratifiable in order to apply our result. The argument we present is entirely constructive and is motivated by the following fact.

**Proposition 3.4.** *Consider a closed, convex cone  $Q \subset \mathbf{R}^n$ , which is neither  $\mathbf{R}^n$  nor a half-space. Then the equality,*

$$Q = \text{coco}(\text{bd } Q).$$

*holds.*

Hence in light of Proposition 3.4, in order to obtain a representation formula for the Clarke subdifferential, it is sufficient to study the (relative) boundary structure of the Clarke normal cone. This is precisely the route we take.<sup>1</sup>

**Lemma 3.5** (Frechét accessibility). *Consider a closed set  $Q \subset \mathbf{R}^n$ , a  $\mathbf{C}^2$ -manifold  $M \subset Q$ , and a point  $\bar{x} \in M$ . Recall that the inclusion  $\hat{N}_Q(\bar{x}) \subset N_M(\bar{x})$  holds. Suppose that a vector  $\bar{v} \in \hat{N}_Q(\bar{x})$  lies in the boundary of  $\hat{N}_Q(\bar{x})$ , relative to the linear space  $N_M(\bar{x})$ . Then there exists a sequence  $(x_i, v_i) \rightarrow (\bar{x}, \bar{v})$ , with  $v_i \in N_Q^P(x_i)$ , and so that all points  $x_i$  lie outside of  $M$ .*

*Proof.* Choose a vector  $\bar{w} \in N_M(\bar{x})$  in such a way so as to guarantee

$$\bar{v} + t\bar{w} \notin \hat{N}_Q(\bar{x}), \quad \text{for all } t > 0.$$

Consider the vectors

$$y(t) := \bar{x} + t(\bar{v} + t\bar{w}), \tag{4}$$

and observe  $y(t) \notin \bar{x} + \hat{N}_Q(\bar{x})$  for every  $t > 0$ . Consider a selection of the projection operator,

$$x(t) \in P_Q(y(t)).$$

Clearly,  $y(t) \rightarrow \bar{x}$  and  $x(t) \rightarrow \bar{x}$ , as  $t \rightarrow 0$ . Observe

$$\begin{aligned} \frac{y(t) - x(t)}{t} &\in N_Q^P(x(t)), \\ x(t) &\neq \bar{x}, \end{aligned} \tag{5}$$

for every  $t$ .

We claim that the points  $x(t)$  all lie outside of  $M$  for all sufficiently small  $t > 0$ . Indeed, if this were not the case, then for sufficiently small  $t$ , the points  $x(t)$  and  $y(t)$  would lie in the prox-normal neighborhood of  $M$  near  $\bar{x}$ , and we would deduce

$$x(t) = P_M(y(t)) = \bar{x},$$

contradicting (5).

Thus all that is left is to show the convergence,  $\frac{y(t) - x(t)}{t} \rightarrow \bar{v}$ . To this end, observe that from (4), we have

$$\frac{y(t) - \bar{x}}{t} \rightarrow \bar{v}. \tag{6}$$

---

<sup>1</sup>The idea to study the boundary structure of the Clarke normal cone in order to establish a convenient representation for the Clarke subdifferential is by no means new. For instance the same idea was used by Rockafellar to establish the representation formula for the convex subdifferential [27, Theorem 25.6]. While working on this paper, the authors became aware that the same strategy was also used to prove a representation formula for the subdifferential of finite-valued, continuous functions whose epigraph has positive reach [12, Theorem 4.9], [13, Theorem 2]. In particular, Proposition 3.4 also appears as [13, Proposition 3].

Hence it suffices to argue  $\frac{x(t)-\bar{x}}{t} \rightarrow 0$ . By definition of  $x(t)$ , we have

$$|y(t) - \bar{x}| \geq |(y(t) - \bar{x}) + (\bar{x} - x(t))|. \quad (7)$$

Squaring and simplifying the inequality above, we obtain

$$\left\langle \frac{y(t) - \bar{x}}{t}, \frac{x(t) - \bar{x}}{t} \right\rangle \geq \frac{1}{2} \left| \frac{x(t) - \bar{x}}{t} \right|^2. \quad (8)$$

From (6) and (7), we deduce that the vectors  $\frac{x(t)-\bar{x}}{t}$  are bounded as  $t \rightarrow 0$ . Consider any limit point  $\gamma \in \mathbf{R}^n$ . Taking the limit in (8), we obtain

$$\langle \bar{v}, \gamma \rangle \geq \frac{1}{2} |\gamma|^2. \quad (9)$$

Since  $\bar{v}$  is a Frechét normal, we deduce

$$\langle \bar{v}, x(t) - \bar{x} \rangle \leq o(|x(t) - \bar{x}|).$$

It immediately follows that

$$\langle \bar{v}, \gamma \rangle \leq 0,$$

and in light of (9), we obtain  $\gamma = 0$ . Hence

$$\frac{y(t) - x(t)}{t} \rightarrow \bar{v},$$

as we claimed.  $\square$

**Remark 3.6.** We note, in passing, that an analogue of Lemma 3.5 (with an identical proof) holds when  $M$  is simply “prox-regular”, in the sense of [24], around  $\bar{x}$ . In particular, the lemma is valid when  $M$  is a convex set.

The combination of Lemma 3.5 and Proposition 3.4 yields dividends even in the simplest case when the manifold  $M$  of Lemma 3.5 is a singleton set. We should emphasize that in the following proposition, we do not even assume that the function in question is directionally Lipschitzian or stratifiable.

**Proposition 3.7** (Isolated singularity). *Consider a continuous function  $f: U \rightarrow \mathbf{R}$ , defined on an open set  $U \subset \mathbf{R}^n$ . Suppose that  $f$  is differentiable on  $U \setminus \{\bar{x}\}$  for some point  $\bar{x} \in U$ , and that  $\partial f(\bar{x}) \neq \emptyset$ . Then*

$$\partial_c f(\bar{x}) = \text{cl} \left( \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \rightarrow \bar{x} \right\} + \text{coco} \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \rightarrow \bar{x} \right\} \right),$$

under the convention that  $\text{conv} \emptyset = \{0\}$ .

*Proof.* Define the two sets

$$E := \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \rightarrow \bar{x} \right\}, \quad H := \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \rightarrow \bar{x} \right\},$$

and consider the epigraph  $Q := \text{epi } f$  and the singleton set  $M := \{(\bar{x}, f(\bar{x}))\}$ .

By Lemma 3.5 and continuity of  $f$ , we have

$$\text{bd } \hat{N}_Q(\bar{x}, f(\bar{x})) \subset \text{coco} (E \times \{-1\}) \cup (H \times \{0\}). \quad (10)$$

CASE 1. Suppose  $\hat{N}_Q(\bar{x}, f(\bar{x}))$  is not equal to  $\mathbf{R}^n \times [0, -\infty)$ . Then from Proposition 3.4 and (10), we deduce

$$N_Q^c(\bar{x}, f(\bar{x})) = \text{cl coco} (E \times \{-1\}) \cup (H \times \{0\}). \quad (11)$$

From (11), we see that an inclusion  $(v, -1) \in N_Q^c(\bar{x}, f(\bar{x}))$  holds if and only if for every  $\epsilon > 0$ , there exist vectors  $y_i \in E \cup H$ , and real numbers  $\lambda_i > 0$ , for  $1 \leq i \leq n+1$ , satisfying

$$\begin{aligned} \left| v - \left( \sum_{i: y_i \in E} \lambda_i y_i + \sum_{i: y_i \in H} \lambda_i y_i \right) \right| &< \epsilon, \\ 1 &= \sum_{i: y_i \in E} \lambda_i \end{aligned}$$

Thus  $\partial_c f(\bar{x}) = \text{cl}(\text{conv } E + \text{coco } H)$ , as we claimed.

CASE 2. Now suppose  $\hat{N}_Q(\bar{x}, f(\bar{x})) = \mathbf{R}^n \times [0, -\infty)$ . Then from (10), we deduce  $H = \mathbf{R}^n$  and  $\partial_c f(\bar{x}) = \mathbf{R}^n = \text{conv } E + \text{coco } H$ , under the convention  $\text{conv } \emptyset = \{0\}$ .  $\square$

As an illustration, consider the following simple example.

**Example 3.8.** Consider the function  $f(x, y) := \sqrt[4]{x^4 + y^2}$  on  $\mathbf{R}^2$ . Clearly  $f$  is differentiable on  $\mathbf{R}^2 \setminus \{(0, 0)\}$ . The gradient has the form

$$\nabla f(x, y) = \frac{1}{(x^4 + y^2)^{3/4}} \begin{pmatrix} x^3 \\ \frac{1}{2}y \end{pmatrix}.$$

From Proposition 3.7, we obtain

$$\partial_c f(\bar{x}) = \text{cl} \left( \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \rightarrow \bar{x} \right\} + \text{coco} \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \rightarrow \bar{x} \right\} \right),$$

Observe that the vectors  $\nabla f(x, 0)$  are equal to  $(\pm 1, 0)$ , and the vectors  $2\sqrt{|y|}\nabla f(0, \pm y)$  are equal to  $(0, \pm 1)$ , whenever  $x \neq 0 \neq y$ . Thus we obtain

$$\begin{aligned} [-1, 1] \times \{0\} &\subset \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \rightarrow \bar{x} \right\}. \\ \{0\} \times \mathbf{R} &\subset \text{coco} \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \rightarrow \bar{x} \right\}. \end{aligned}$$

Consequently, the inclusion

$$[-1, 1] \times \mathbf{R} \subset \partial_c f(0, 0)$$

holds. The absolute value of the first coordinate of  $\nabla f(x, y)$  is always bounded by 1, which implies the reverse inclusion above. Thus we have exact equality  $\partial_c f(0, 0) = [-1, 1] \times \mathbf{R}$ .

We record the following observation for ease of reference.

**Corollary 3.9.** *Consider a closed, convex cone  $Q \subset \mathbf{R}^n$  with nonempty interior. Suppose that  $\text{bd } Q$  is contained in a proper linear subspace. Then  $Q$  is either all of  $\mathbf{R}^n$  or a half-space.*

*Proof.* Clearly if  $Q$  were neither  $\mathbf{R}^n$  or a half-space, then by Proposition 3.4, we would deduce that  $Q = \text{coco}(\text{bd } Q)$  has empty interior, which is a contradiction.  $\square$

Armed with Proposition 3.4 and Lemma 3.5, we can now prove the main result of this section with ease.

**Theorem 3.10** (Characterization). *Consider a proper, stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is finite at  $\bar{x}$ . Suppose that  $f$  is vertically continuous and directionally Lipschitzian at all points of  $\text{dom } f$  near  $\bar{x}$ . Then for any dense subset  $\Omega \subset \text{dom } f$ , we have*

$$N_{\text{epi } f}^c(\bar{x}, f(\bar{x})) = \text{coco} \left\{ \lim_{i \rightarrow \infty} \frac{(\nabla f(x_i), -1)}{\sqrt{1 + |\nabla f(x_i)|}} : x_i \xrightarrow[f]{\Omega} \bar{x} \right\} + (N_{\text{dom } f}^c(\bar{x}) \times \{0\}). \quad (12)$$

Consequently, the Clarke subdifferential admits the presentation

$$\partial_c f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\} + \text{coco} \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\} + N_{\text{dom } f}^c(\bar{x}).$$

*Proof.* We first prove (12). Observe that since  $f$  is vertically continuous at  $\bar{x}$ , we have  $N_{\text{dom } f}^c(\bar{x}) \times \{0\} \subset N_{\text{epi } f}^c(\bar{x}, f(\bar{x}))$ , and hence the inclusion “ $\supset$ ” holds. Therefore we must establish the reverse inclusion. To this effect, intersecting the domain of  $f$  with a small open ball around  $\bar{x}$ , we may assume that  $f$  is directionally Lipschitzian and vertically continuous at each point  $x \in \text{dom } f$ . For notational convenience, for a vector  $v \in \mathbf{R}^n$ , let  $\bar{v} := \frac{(v, -1)}{\sqrt{1+|v|^2}}$ . Define the set-valued mapping

$$F(x) := \text{coco} \left( \overline{\left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} x \right\}} \cup (N_{\text{dom } f}(x) \cap \mathbf{B}) \times \{0\} \right), \quad (13)$$

By Proposition 3.3, the set  $\overline{\left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} x \right\}}$  is nonempty. Furthermore, from the established inclusion “ $\supset$ ”, we see that  $N_{\text{dom } f}(x)$  is pointed and hence the set  $\text{coco } N_{\text{dom } f}(x)$  is closed for all  $x \in \text{dom } f$ . Consequently, we deduce

$$F(x) = \text{coco} \left\{ \overline{\left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} x \right\}} + (N_{\text{dom } f}^c(x) \times \{0\}) \right\},$$

Combining (13) with Lemma 2.2, we see that  $F$  is outer-semicontinuous with respect to  $f$ -attentive convergence.

Now consider a stratification of  $\text{dom } f$  into manifolds  $\{M_i\}$  having the property that  $f$  is smooth on each stratum  $M_i$ . Restricting the domain of  $f$ , we may assume that the stratification  $\{M_i\}$  consists of only finitely many sets. We prove the theorem by induction on the dimension of the strata  $M_i$  in which the point  $\bar{x}$  lies.

Clearly, the result holds for all strata of dimension  $n$ , since  $f$  is smooth on such strata and  $\Omega$  is dense in  $\text{dom } f$ . As an inductive hypothesis, suppose that the claim holds for all strata that are strictly greater in the partial order  $\preceq$  than a certain stratum  $M$  and let  $\bar{x}$  be an arbitrary point of  $M$ .

Since  $f$  is smooth on  $M$ , we deduce that  $\text{gph } f|_M$  is a smooth manifold. Then by Lemma 3.5, for every vector  $0 \neq v \in \text{rb } \hat{N}_{\text{epi } f}(\bar{x}, f(\bar{x}))$ , there exists a sequence  $(x_l, r_l, v_l) \rightarrow (\bar{x}, f(\bar{x}), v)$ , with  $v_l \in \hat{N}_{\text{epi } f}(x_l, r_l)$  and  $(x_l, r_l) \notin \text{gph } f|_M$ . Suppose that there exists a subsequence satisfying  $r_l \neq f(x_l)$  for each index  $l$ . Then since  $f$  is vertically continuous at  $\bar{x}$ , we obtain

$$v = \lim_{l \rightarrow \infty} v_l \subset \limsup_{l \rightarrow \infty} N_{\text{epi } f}(x_l, r_l) \subset N_{\text{dom } f}(\bar{x}) \times \{0\} \subset F(\bar{x}).$$

On the other hand, if  $r_l = f(x_l)$  for all large indices  $l$ , then restricting to a subsequence, we may assume that all the points  $x_l$  lie in a stratum  $M'$  with  $M' \succ M$ . The inductive hypothesis and  $f$ -attentive outer-semicontinuity of  $F$  yield the inclusion  $v \in F(\bar{x})$ .

Thus we have established the inclusion,

$$\text{rb } \hat{N}_{\text{epi } f}(\bar{x}, f(\bar{x})) \subset F(\bar{x}).$$

Since  $\partial^\infty f(\bar{x})$  is pointed, we deduce that the cone  $\hat{N}_{\text{epi } f}(\bar{x})$  is neither a linear subspace nor a half-subspace. Consequently by Lemma 2.2, we deduce

$$\hat{N}_{\text{epi } f}(\bar{x}, f(\bar{x})) = \text{coco } \text{rb } \hat{N}_{\text{epi } f}(\bar{x}, f(\bar{x})) \subset F(\bar{x}). \quad (14)$$

In fact, we have shown that (14) holds for all points  $\bar{x} \in M$ .

Finally consider a limiting normal  $v \in N_{\text{epi } f}(\bar{x})$ . Then there exists a sequence  $(x_l, r_l, v_l) \rightarrow (\bar{x}, f(\bar{x}), v)$ , with  $v_l \in \hat{N}_{\text{epi } f}(x_l, r_l)$ . It follows from (14), the inductive hypothesis, and  $f$ -attentive outer-semicontinuity of  $F$  that the inclusion  $v \in F(\bar{x})$  holds. Thus the induction is complete, as is the proof of (12).

To finish the proof of the theorem, define the two sets

$$E := \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\}, \quad H := \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\}. \quad (15)$$

Observe

$$\text{coco} \left\{ \overline{\left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\}} \right\} = \text{coco} (E \times \{-1\}) \cup (H \times \{0\}).$$

Thus an inclusion  $(v, -1) \in N_Q^c(\bar{x}, f(\bar{x}))$  holds if and only if there exist vectors  $y_i \in E \cup H$  and  $y \in N_{\text{dom } f}^c(\bar{x})$ , and real numbers  $\lambda_i > 0$ , for  $1 \leq i \leq n+1$ , satisfying

$$\begin{aligned} v &= \sum_{i: y_i \in E} \lambda_i y_i + \sum_{i: y_i \in H} \lambda_i y_i + y, \\ 1 &= \sum_{i: y_i \in E} \lambda_i \end{aligned}$$

The result follows.  $\square$

Recovering representation (ACLR) of the introduction, in the setting of stratifiable functions, is now an easy task.

**Corollary 3.11.** *Consider a proper, stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ . Suppose that  $f$  is directionally Lipschitzian at all points of  $\text{dom } f$  near  $\bar{x}$ , and is continuous near  $\bar{x}$  relative to the domain of  $f$ . Then we have*

$$\partial_c f(\bar{x}) = \bigcap_{\delta > 0} \text{cl conv} \left( \nabla f(\Omega \cap B_\delta(\bar{x})) \right) + N_{\text{dom } f}^c(\bar{x}),$$

where  $\Omega$  is any dense subset of  $\text{dom } f$ .

*Proof.* Since the cone  $N_{\text{epi } f}^c(\bar{x}, f(\bar{x}))$  is pointed, one can easily verify, much along the lines of Lemma 2.2, that the equation

$$\bigcap_{\delta > 0} \text{cl coco} \left\{ \frac{(\nabla f(x), -1)}{\sqrt{1 + |\nabla f(x)|}} : x \in \Omega \cap B_\delta(\bar{x}) \right\} = \text{coco} \left\{ \lim_{i \rightarrow \infty} \frac{(\nabla f(x_i), -1)}{\sqrt{1 + |\nabla f(x_i)|}} : x_i \xrightarrow{\Omega} \bar{x} \right\},$$

holds. The result follows by an application of Theorem 3.10. We leave the details to the reader.  $\square$

Our next goal is to recover the representation of the convex subdifferential (CoR) of the introduction in the stratifiable setting. In fact, we will consider the more general case of amenable functions. Before proceeding, recall that a proper, lower-semicontinuous, convex function is directionally Lipschitzian at some point if and only if its domain has nonempty interior. A completely analogous situation occurs for amenable functions.

**Lemma 3.12.** *Consider a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is amenable at  $\bar{x}$ . Let  $V$  be a neighborhood of  $\bar{x}$  so that  $f$  can be written as a composition  $f = g \circ F$ , for a  $\mathbf{C}^1$  mapping  $F: V \rightarrow \mathbf{R}^m$  and a proper, lower-semicontinuous, convex function  $g: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$ , so that the qualification condition*

$$N_{\text{dom } g}(F(\bar{x})) \cap \ker \nabla F(\bar{x})^* = \{0\},$$

*is satisfied. Then there exists a neighborhood  $U$  of  $\bar{x}$  so that*

1.  $F(U \cap \text{int dom } f) \subset \text{int dom } g$ ,
2.  $U \cap F^{-1}(\text{int dom } g) \subset \text{int dom } f$ .

*Furthermore  $f$  is directionally Lipschitzian at  $\bar{x}$  if and only if  $\bar{x}$  lies in  $\text{cl}(\text{int dom } f)$ .*

*Proof.* Let us first recall a few useful formulas. To this end, [23, Theorem 3.3] shows that there exists a neighborhood  $U$  of  $\bar{x}$  so that for all points  $x \in U \cap \text{dom } f$ , we have

$$\{0\} = N_{\text{dom } g}(F(x)) \cap \ker \nabla F(x)^*, \quad (16)$$

$$\partial f(x) = \nabla F(x)^* \partial g(F(x)), \quad (17)$$

$$N_{\text{dom } f}(x) = \nabla F(x)^* N_{\text{dom } g}(F(x)). \quad (18)$$

Furthermore a computation in the proof of Proposition 2.14 (item 5) shows that for any  $x \in U \cap \text{dom } f$  and any  $r > f(x)$ , we have

$$N_{\text{epi } f}(x, r) = \nabla F(x)^* N_{\text{dom } g}(F(x)) \times \{0\}. \quad (19)$$

Observe for any  $x \in U \cap \text{int dom } f$ , we have

$$0 = N_{\text{dom } f}(x) = \nabla F(x)^* N_{\text{dom } g}(F(x)),$$

and consequently  $N_{\text{dom } g}(F(x)) = 0$ . We conclude  $F(x) \in \text{int dom } g$ , thus establishing 1.

Now consider a point  $x \in U \cap F^{-1}(\text{int dom } g)$ . Using (19), we deduce  $N_{\text{epi } f}(x, r) = 0$  for any  $r > f(x)$ . Hence by [29, Exercise 6.19], we conclude  $(x, r) \in \text{int epi } f$  and consequently  $x \in \text{int dom } f$ , thus establishing 2.

By [29, Exercise 10.25 (a)], we have  $\partial^\infty f(\bar{x}) = N_{\text{dom } f}(\bar{x})$ , and in light of (16) and (18) one can readily verify that the cone  $N_{\text{dom } f}(\bar{x})$  is pointed if and only if  $N_{\text{dom } g}(F(\bar{x}))$  is pointed.

Now suppose that  $\bar{x}$  lies in  $\text{cl}(\text{int dom } f)$ . Then by 1 the domain  $\text{dom } g$  has nonempty interior and consequently  $N_{\text{dom } g}(F(\bar{x}))$  is pointed, as is  $N_{\text{dom } f}(\bar{x})$ .

Conversely suppose that  $f$  is directionally Lipschitzian at  $\bar{x}$ . Then  $N_{\text{dom } g}(F(\bar{x}))$  is pointed, and consequently  $\text{dom } g$  has nonempty interior. Observe since  $F$  is continuous, the set  $F^{-1}(\text{int dom } g) \subset \text{dom } f$  is open. Hence it is sufficient to argue that this set contains  $\bar{x}$  in its closure. Suppose this is not the case. Then there exists a neighborhood  $U$  of  $\bar{x}$  so that the image  $F(U)$  does not intersect  $\text{int dom } g$ . It follows that the range of the linearised mapping  $w \mapsto F(\bar{x}) + \nabla F(\bar{x})w$  can be separated from  $\text{dom } g$ , thus contradicting (16). See [29, Theorem 10.6] for a more detailed explanation of this latter assertion.  $\square$

We can now easily recover, in the stratifiable setting, representation (CoR) of the introduction. In fact, an entirely analogous formula holds more generally for amenable functions.

**Corollary 3.13.** *Consider a proper stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is amenable at a point  $\bar{x}$ , and so that  $\bar{x}$  lies in the closure of the interior of  $\text{dom } f$ . Let  $\Omega$  be any dense subset of  $\text{dom } f$ . Then the subdifferential admits the presentation*

$$\partial f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\} + N_{\text{dom } f}(\bar{x}).$$

*Proof.* By [29, Exercise 10.25], we have  $\partial^\infty f(\bar{x}) = N_{\text{dom } f}(\bar{x})$ . Thus we have

$$\text{coco} \left\{ \lim_{\substack{i \rightarrow \infty \\ t_i \downarrow 0}} t_i \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\} \subset N_{\text{dom } f}(\bar{x}).$$

Observe  $f$  is amenable, directionally Lipschitzian (Lemma 3.12), and vertically continuous (Proposition 2.14) at each point of  $\text{dom } f$  near  $\bar{x}$ . Applying Theorem 3.10, we deduce

$$\partial f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow[f]{\Omega} \bar{x} \right\} + N_{\text{dom } f}(\bar{x}).$$

Noting that the subdifferential map  $\partial f$  of an amenable function is outer-semicontinuous, the result follows.  $\square$

A natural question arises. Does the corollary above hold more generally without the stratifiability assumption? The answer turns out to be yes. This is immediate, in light of (CoR), for the subclass of *lower- $\mathbf{C}^2$*  functions (those functions that are locally representable as a difference of convex functions and convex quadratics). A first attempt at a proof for general amenable functions might be to consider the representation  $f = g \circ F$  and the chain rule

$$\partial f(x) = \nabla F(\bar{x})^* \partial g(F(\bar{x})).$$

One may then try to naively use Rockafellar's representation formula (CoR) for the convex subdifferential

$$\partial g(F(\bar{x})) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla g(y_i) : y_i \rightarrow F(\bar{x}) \right\}$$

to deduce the result. However, we immediately run into trouble since  $F$  may easily fail to be surjective onto a neighborhood of  $F(\bar{x})$  in  $\text{dom } g$ . Hence a different more sophisticated proof technique is required. For completeness, we present an argument below, which is a natural extension of the proof of [27, Theorem 25.6]. It is furthermore instructive to emphasize how the stratifiability assumption allowed us in Corollary 3.13 to bypass essentially all the technical details of the argument below.

**Theorem 3.14.** Consider a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is amenable at a point  $\bar{x}$  lying in  $\text{cl}(\text{int dom } f)$ . Then the subdifferential admits the presentation

$$\partial f(\bar{x}) = \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\} + N_{\text{dom } f}(\bar{x}),$$

where  $\Omega$  is any full-measure subset of  $\text{dom } f$ .

*Proof.* Recall that  $f$  is Clarke regular at  $\bar{x}$ , and therefore  $\partial^\infty f(\bar{x}) = N_{\text{dom } f}(\bar{x})$  is the recession cone of  $\partial f(\bar{x})$ . Combining this with the fact that the map  $\partial f$  is outer-semicontinuous at  $\bar{x}$ , we immediately deduce the inclusion “ $\supset$ ”.

We now argue the reverse inclusion. To this end, let  $V$  be a neighborhood of  $\bar{x}$  so that  $f$  can be written as a composition  $f = g \circ F$ , for a  $\mathbf{C}^1$  mapping  $F: V \rightarrow \mathbf{R}^m$  and a proper, lower-semicontinuous, convex function  $g: \mathbf{R}^m \rightarrow \overline{\mathbf{R}}$ , so that the qualification condition

$$N_{\text{dom } g}(F(\bar{x})) \cap \ker \nabla F(\bar{x})^* = \{0\},$$

is satisfied. Since  $f$  is directionally Lipschitzian at  $\bar{x}$ , the subdifferential  $\partial f(x)$  is the sum of the convex hull of its extreme points and the recession cone  $N_{\text{dom } f}(\bar{x})$ . Furthermore every extreme point is a limit of exposed points. Thus

$$\partial f(\bar{x}) = \text{conv}(\text{cl } E) + N_{\text{dom } f}(\bar{x}),$$

where  $E$  is the set of all exposed point of  $\partial f(\bar{x})$ .

Hence to prove the needed inclusion, it suffices to argue the inclusion

$$E \subset \text{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\}.$$

Here, we should note that since  $f$  is directionally Lipschitzian at  $\bar{x}$ , the set on the right hand side is closed.

To this end, let  $\bar{v}$  be an arbitrary exposed point of  $\partial f(\bar{x})$ . By definition, there exists a vector  $\bar{a} \in \mathbf{R}^n$  with  $|\bar{a}| = 1$  and satisfying

$$\langle \bar{a}, \bar{v} \rangle > \langle \bar{a}, v \rangle \text{ for all } v \in \partial f(\bar{x}) \text{ with } v \neq \bar{v}.$$

Since  $N_{\text{dom } f}(\bar{x})$  is the recession cone of  $\partial f(\bar{x})$ , from above we deduce

$$\langle \bar{a}, z \rangle < 0 \text{ for all } 0 \neq z \in N_{\text{dom } f}(\bar{x}),$$

and consequently

$$\langle \nabla F(\bar{x})\bar{a}, w \rangle < 0 \text{ for all } 0 \neq w \in N_{\text{dom } g}(F(\bar{x})).$$

Consider the half-line  $\{F(\bar{x}) + t\nabla F(\bar{x})\bar{a} : t \geq 0\}$ . We claim that this half-line cannot be separated from  $\text{dom } g$ . Indeed, otherwise there would exist a nonzero vector  $\bar{w} \in N_{\text{dom } g}(F(\bar{x}))$  so that for all  $t > 0$  and all  $x \in \text{dom } g$  we have

$$\langle x, \bar{w} \rangle \leq \langle F(\bar{x}) + t\nabla F(\bar{x})\bar{a}, \bar{w} \rangle < \langle F(\bar{x}), \bar{w} \rangle,$$

which is a contradiction. Hence by [27, Theorem 11.3], this half-line must meet the interior of  $\text{dom } g$ . By convexity then there exists a real number  $\alpha > 0$  satisfying

$$\{F(\bar{x}) + t\nabla F(\bar{x})\bar{a} : 0 < t \leq \alpha\} \subset \text{int}(\text{dom } g).$$

Consequently the points  $F(\bar{x} + t\bar{a})$  lie in  $\text{int dom } g$  for all sufficiently small  $t > 0$ . By Lemma 3.12, we deduce that there exists a real number  $\beta > 0$  so that

$$\{\bar{x} + t\bar{a} : 0 < t \leq \beta\} \subset \text{int}(\text{dom } f).$$

Hence  $f$  is Lipschitz continuous at each point  $\bar{x} + t\bar{a}$  (for  $0 < t \leq \beta$ ), and so from (LipR) we obtain

$$\partial f(\bar{x} + t\bar{a}) = \text{conv} \left\{ \lim_{j \rightarrow \infty} \nabla f(x_j) : x_j \xrightarrow{\Omega} \bar{x} + t\bar{a} \right\}. \quad (20)$$



Now choose a sequence  $t_i \rightarrow 0$  and observe that by [27, Theorem 24.6], for any  $\epsilon > 0$  we have

$$\begin{aligned}\partial g(F(\bar{x} + t_i \bar{a})) &\subset \operatorname{argmax}_{v \in \partial g(F(\bar{x}))} \langle \nabla F(\bar{x}) \bar{a}, v \rangle + \epsilon \mathbf{B}, \\ &= \operatorname{argmax}_{v \in \partial g(F(\bar{x}))} \langle \bar{a}, \nabla F(\bar{x})^* v \rangle + \epsilon \mathbf{B},\end{aligned}$$

for all large  $i$ . We deduce,

$$\nabla F(\bar{x})^* \partial g(F(\bar{x} + t_i \bar{a})) \subset \operatorname{argmax}_{w \in \partial f(\bar{x})} \langle \bar{a}, w \rangle + \epsilon \mathbf{B} = \bar{v} + \epsilon \mathbf{B}.$$

Thus there exists a sequence  $w_i \in \partial g(F(\bar{x} + t_i \bar{a}))$  with  $\nabla F(\bar{x})^* w_i \rightarrow \bar{v}$ . Consequently the vectors  $\nabla F(\bar{x} + t_i \bar{a})^* w_i \in \partial f(\bar{x} + t_i \bar{a})$  converge to  $\bar{v}$ . The result now easily follows from (20) and the fact that  $f$  is directionally Lipschitzian at  $\bar{x}$ .  $\square$

The following is a further illustration of the applicability of our results to a wide variety of situations.

**Corollary 3.15.** *Consider a proper stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  that is locally Lipschitz continuous at a point  $\bar{x}$ , relative to  $\operatorname{dom} f$ . Suppose furthermore that  $\operatorname{dom} f$  is an epi-Lipschitzian set at  $\bar{x}$ . Then the formula*

$$\partial f(\bar{x}) = \operatorname{conv} \left\{ \lim_{i \rightarrow \infty} \nabla f(x_i) : x_i \xrightarrow{\Omega} \bar{x} \right\} + N_{\operatorname{dom} f}^c(\bar{x}),$$

holds, where  $\Omega$  is any dense subset of  $\operatorname{dom} f$ .

*Proof.* Since  $f$  is locally Lipschitz near  $\bar{x}$  relative to  $\operatorname{dom} f$ , there exists a globally Lipschitz function  $\tilde{f}: \mathbf{R}^n \rightarrow \mathbf{R}$ , agreeing with  $f$  on  $\operatorname{dom} f$  near  $\bar{x}$ . Hence, we have

$$f(x) = \tilde{f}(x) + \delta_{\operatorname{dom} f}(x), \text{ locally near } \bar{x}.$$

Combining this with [29, Exercise 10.10], we deduce

$$\partial^\infty f(x) \subset N_{\operatorname{dom} f}(x), \text{ for } x \text{ near } \bar{x}.$$

We conclude that  $f$  is directionally Lipschitzian at all points of  $\operatorname{dom} f$  near  $\bar{x}$ , and furthermore since the gradients of  $\tilde{f}$  are bounded near  $\bar{x}$  so are the gradients of  $f$ . The result follows immediately by an application of Proposition 2.14 and Theorem 3.10.  $\square$

## 4 Comments on random sampling

Consider a proper, stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ , directionally Lipschitzian at all points of  $\operatorname{dom} f$  near  $\bar{x}$ , and is continuous near  $\bar{x}$  relative to the domain of  $f$ . Let us suppose for the time being that the normal cone  $N_{\operatorname{dom} f}(\bar{x})$  is known exactly. Our interest lies in approximating the Clarke subdifferential  $\partial_c f(\bar{x})$ . In light of Corollary 3.11, one can quickly modify the method considered in [3], in order to achieve this goal. We now describe this procedure in more detail.

Fix a certain radius  $\delta > 0$  and a sample space  $\Lambda = B_\delta(\bar{x})$ , along with an associated probability measure that is absolutely continuous with respect to the Lebesgue measure  $\mu$  on  $\mathbf{R}^n$ . We assume that the corresponding density  $\theta$  is strictly positive almost everywhere on  $\Lambda$ . Observe that the set  $\Lambda \cap \operatorname{dom} \nabla f$  has nonempty interior, and consequently has positive probability. We can consider a sequence of independent trials with outcomes  $x_i \in \Lambda$  for  $i = 1, 2, \dots$ , and form trial sets

$$C_k = \operatorname{conv} \{ \nabla f(x_i) : x_i \in \Lambda \cap \operatorname{dom} \nabla f, 1 \leq i \leq k \}.$$

One may then hope that the sets  $D_k := C_k + N_{\operatorname{dom} f}^c(\bar{x})$  approximate the Clarke subdifferential  $\partial_c f(\bar{x})$  well, as  $k$  tends to infinity. One can verify this rigorously via some modifications of arguments made in [3]. The starting point is the following result. We omit the proof since it is identical to the argument in [3, Theorem 2.1].

**Theorem 4.1** (Limiting Approximation). *Consider a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  and a point  $\bar{x} \in \text{dom } f$ . Suppose that  $f$  is continuously differentiable on an open set dense in  $\text{dom } f$ . Then for any sampling radius  $\delta$ , we have*

$$\text{cl} \bigcup_{k=1}^{\infty} C_k = \text{cl conv } \nabla f(B_{\delta}(\bar{x})), \text{ almost surely.}$$

The following theorem establishes that  $\text{cl} \bigcup_{k=1}^{\infty} D_k$  is almost surely an outer approximation of  $\partial_c f(\bar{x})$ .

**Theorem 4.2** (Outer approximation). *Consider a proper, stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ . Suppose that  $f$  is directionally Lipschitzian at all points of  $\text{dom } f$  near  $\bar{x}$ , and is continuous near  $\bar{x}$  relative to the domain of  $f$ . Then*

$$\partial_c f(\bar{x}) \subset \text{cl} \bigcup_{k=1}^{\infty} D_k, \text{ almost surely.}$$

*Proof.* This is immediate from Theorem 4.1 and Corollary 3.11.  $\square$

On the other hand, the following lemma shows that  $\text{cl} \bigcup_{k=1}^{\infty} D_k$  is not too much bigger than  $\partial_c f(\bar{x})$ , as long as the radius is sufficiently small and we restrict ourselves to considering only bounded subsets of  $\mathbf{R}^n$ .

**Lemma 4.3** (Truncated Inner approximation). *Consider a proper, stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ . Suppose that  $f$  is directionally Lipschitzian at all points of  $\text{dom } f$  near  $\bar{x}$ , and is continuous near  $\bar{x}$  relative to the domain of  $f$ . Then for any compact subset  $\Gamma \subset \mathbf{R}^n$  and a real number  $\epsilon > 0$  we have, for any sufficiently small sampling radius,*

$$\Gamma \cap \text{cl} \bigcup_{k=1}^{\infty} D_k \subset \Gamma \cap \partial_c f(\bar{x}) + \epsilon \mathbf{B}, \text{ almost surely.}$$

*If, in addition,  $f$  is Lipschitz continuous on its domain, then for any real number  $\epsilon > 0$  we have, for any sufficiently small sampling radius,*

$$\text{cl} \bigcup_{k=1}^{\infty} D_k \subset \partial_c f(\bar{x}) + \epsilon \mathbf{B}, \text{ almost surely.}$$

*Proof.* Since  $f$  is directionally Lipschitzian at  $\bar{x}$ , one can easily see check that the mapping  $\partial_c f$  is outer-semicontinuous at  $\bar{x}$ . Consequently the truncated mapping  $x \mapsto \Gamma \cap \partial_c f(x)$  is upper-semicontinuous at  $\bar{x}$ . We conclude that there exists a real number  $\delta > 0$  such that

$$\Gamma \cap \partial_c f(\bar{x} + \delta \mathbf{B}) \subset \Gamma \cap \partial_c f(\bar{x}) + \epsilon \mathbf{B}.$$

Observe that  $\text{cl} \bigcup_{k=1}^{\infty} D_k$  is almost surely contained in  $\Gamma \cap \text{cl conv } \partial_c f(\bar{x} + \delta \mathbf{B})$ , and hence the result follows.

Now suppose that  $f$  is Lipschitz continuous on its domain. Then, one can verify that that there exists a real number  $\delta > 0$  satisfying

$$\nabla f(B_{\delta}(\bar{x})) \subset \partial_c f(\bar{x}) + \epsilon \mathbf{B}.$$

Since  $\text{cl} \bigcup_{k=1}^{\infty} D_k$  is almost surely contained  $\text{cl conv}(\nabla f(B_{\delta}(\bar{x})) + N_{\text{dom } f}(\bar{x}))$ , the result follows.  $\square$

In particular, the distance of a fixed vector  $v$  to  $\text{cl} \bigcup_{k=1}^{\infty} D_k$  can be made arbitrarily close to the distance of  $v$  to  $\partial_c f(\bar{x})$ .

**Theorem 4.4** (Distance approximation). *Consider a proper, stratifiable function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ . Suppose that  $f$  is directionally Lipschitzian at all points of  $\text{dom } f$  near  $\bar{x}$ , and is continuous near  $\bar{x}$  relative to the domain of  $f$ . Then for any vector  $v \in \mathbf{R}^n$  and a real  $\epsilon > 0$  we have, for any sufficiently small sampling radius,*

$$\text{dist}(v, \text{cl} \bigcup_{k=1}^{\infty} D_k) \geq \text{dist}(v, \partial_c f(\bar{x})) - \epsilon, \text{ almost surely,}$$

and consequently

$$\lim_{k \rightarrow \infty} |\text{dist}(v, \text{cl } D_k) - \text{dist}(v, \partial_c f(\bar{x}))| < \epsilon, \text{ almost surely.}$$

*Proof.* Let  $\gamma := \text{dist}(v, \partial_c f(\bar{x}))$  and  $\Gamma = \overline{B}_\gamma(v)$ . Then by Lemma 4.3, we have for any sufficiently small radius,

$$\Gamma \cap \text{cl } \bigcup_{k=1}^{\infty} D_k \subset \Gamma \cap \partial_c f(\bar{x}) + \epsilon \mathbf{B}, \text{ almost surely.}$$

The result readily follows from the inclusion above and Theorem 4.2.  $\square$

A natural test for optimality, using the sampling scheme, is to determine whether the inclusion

$$0 \in D_k,$$

holds. According to Theorem 4.2, if  $\bar{x}$  is a Clarke critical point, then  $\text{dist}(0, D_k) \rightarrow 0$  almost surely. On the other hand, if  $\bar{x}$  is not Clarke-critical, then by Theorem 4.4, for any sufficiently small radius, we have  $\lim_{k \rightarrow \infty} \text{dist}(0, \text{cl } D_k) > 0$ , and hence the test  $0 \in D_k$  will not generate a false positive.

From a computational point of view, there is a certain difficulty we have not addressed. Suppose that for each radius  $\delta$ , the trial points  $x_i$  are sampled with uniform distribution on the ball  $B_\delta(\bar{x})$ . One could worry that as the sampling radius decreases to zero, the proportion of points  $x_i$  discarded (those that lie outside of the domain) to those that are in the domain could be arbitrarily large, with positive probability. This, for instance, could happen if the domain was the set  $\{(x, y) \in \mathbf{R}^2 : |y| \leq x^2, 0 \leq x\}$ . This pathology, however, does not occur in the directionally Lipschitzian setting.

**Proposition 4.5** (Domain of a continuous directionally Lipschitzian function). *Consider a proper function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , that is finite at  $\bar{x}$ . Suppose that  $f$  is directionally Lipschitzian at  $\bar{x}$  and is continuous at  $\bar{x}$ , relative to  $\text{dom } f$ . Then the domain of  $f$  is epi-Lipschitzian at  $\bar{x}$ .*

*Proof.* By continuity of  $f$ , the domain of  $f$  is locally closed near  $\bar{x}$ . Now observe that for each real number  $r$  satisfying  $r > f(\bar{x})$ , we have  $N_{\text{epi } f}(\bar{x}, r) = N_{\text{dom } f}(\bar{x}) \times \{0\}$ . Consequently the inclusion

$$N_{\text{dom } f}(\bar{x}) \times \{0\} \subset N_{\text{epi } f}(\bar{x}, f(\bar{x})),$$

holds. We conclude that the normal cone  $N_{\text{dom } f}(\bar{x})$  is pointed, and hence the domain of  $f$  is an epi-Lipschitzian set at  $\bar{x}$ .  $\square$

The *lower Lebesgue density* of a set  $Q \subset \mathbf{R}^n$  at a point  $\bar{x}$  is

$$\text{Dens}^-(Q, \bar{x}) := \liminf_{\delta \rightarrow 0} \frac{\mu(Q \cap B_\delta(\bar{x}))}{\mu(B_\delta(\bar{x}))},$$

where we recall that  $\mu$  denotes the Lebesgue measure on  $\mathbf{R}^n$ .

**Proposition 4.6** (Density of epi-Lipschitzian sets). *Consider a set  $Q \subset \mathbf{R}^n$  that is epi-Lipschitzian at a point  $\bar{x}$ . Then we have*

$$\text{Dens}^-(Q, \bar{x}) > 0.$$

*Proof.* Since  $Q$  is epi-Lipschitzian at  $\bar{x}$ , we may assume that  $Q$  is an epigraph of a Lipschitz continuous function  $f: \mathbf{R}^{n-1} \rightarrow \mathbf{R}$ , with  $\bar{x} = (0, 0)$  (See for example [28, Section 4]). We deduce  $f(x) \leq \kappa|x|$ , for some constant  $\kappa > 0$  and for all points in  $x \in \mathbf{R}^{n-1}$ . Consequently,  $\text{epi } f$  contains a convex cone with nonempty interior, a set that has strictly positive Lebesgue density. The result follows.  $\square$

The sampling scheme outlined in this section is effective whenever gradients are cheap to compute and the normal cone to the domain at the point of interest is known in advance. Now suppose that the normal cone is unknown to us, but nevertheless we can test whether a point is in the domain and we can project onto the domain easily. Then a slight modification of the sampling scheme can still be effective at approximating the Clarke subdifferential. Namely during our sampling scheme, rather than discarding points lying outside of the domain, we may use these points to approximate the normal cone by projecting onto the domain. See [15] for more details. Using this information along with the sampled gradients, we may still hope to approximate the whole Clarke subdifferential effectively. We, however, do not pursue this further in our work.

## 5 Local Dimension of semi-algebraic subdifferential graphs.

In this section, we apply our methods to study the size of subdifferential graphs of semi-algebraic functions. We first establish notation and record some preliminary facts from semi-algebraic geometry.

### 5.1 Semi-algebraic Geometry.

A *semi-algebraic* set  $S \subset \mathbf{R}^n$  is a finite union of sets of the form

$$\{x \in \mathbf{R}^n : P_1(x) = 0, \dots, P_k(x) = 0, Q_1(x) < 0, \dots, Q_l(x) < 0\},$$

where  $P_1, \dots, P_k$  and  $Q_1, \dots, Q_l$  are polynomials in  $n$  variables. In other words,  $S$  is a union of finitely many sets, each defined by finitely many polynomial equalities and inequalities. A map  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  is said to be *semi-algebraic* if  $\text{gph } F \subset \mathbf{R}^{n+m}$  is a semi-algebraic set. Semi-algebraic sets enjoy many nice structural properties. Unless otherwise stated, we follow the notation of [8] and [30].

A fundamental fact about semi-algebraic sets is provided by the Tarski-Seidenberg Theorem [8, Theorem 2.3]. Roughly speaking, it states that a linear projection of a semi-algebraic set remains semi-algebraic. From this result, it follows that a great many constructions preserve semi-algebraicity. In particular, for a semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , the set-valued mappings  $\partial_P f$ ,  $\hat{\partial} f$ ,  $\partial f$ , and  $\partial_c f$  are semi-algebraic. See for example [18, Proposition 3.1].

**Definition 5.1** (Compatibility). Given finite collections  $\{B_i\}$  and  $\{C_j\}$  of subsets of  $\mathbf{R}^n$ , we say that  $\{B_i\}$  is *compatible* with  $\{C_j\}$  if for all  $B_i$  and  $C_j$ , either  $B_i \cap C_j = \emptyset$  or  $B_i \subset C_j$ .

**Definition 5.2** (Stratifications). Consider a set  $Q$  in  $\mathbf{R}^n$ . A *stratification* of  $Q$  is a finite partition of  $Q$  into disjoint, connected, manifolds  $M_i$  (called strata) with the property that for each index  $i$ , the intersection of the closure of  $M_i$  with  $Q$  is the union of some  $M_j$ 's.

Remarkably, semi-algebraic sets always admit stratifications. Indeed, the following stronger result holds.

**Proposition 5.3.** [30, Theorem 4.8] *Consider a semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . Then there exists a stratification  $\mathcal{A}$  of  $\text{dom } f$  so that  $f$  is smooth on each stratum  $M_i \in \mathcal{A}$ . Furthermore, if  $\mathcal{B}$  is some other stratification of  $\text{dom } f$ , then we can ensure that  $\mathcal{A}$  is compatible with  $\mathcal{B}$ .*

**Definition 5.4** (Dimension). Let  $Q \subset \mathbf{R}^n$  be a nonempty semi-algebraic set. Then the *dimension* of  $Q$ , denoted  $\dim Q$ , is the maximal dimension of a stratum in any stratification of  $Q$ . We adopt the convention that  $\dim \emptyset = -\infty$ .

It can be easily shown that the dimension does not depend on the particular stratification. Observe that the dimension of a semi-algebraic set only depends on the maximal dimensional manifold in a stratification. Hence, dimension is a crude measure of the size of the semi-algebraic set. This motivates a localized notion of dimension.

**Definition 5.5** (Local dimension). Consider a semi-algebraic set  $Q \subset \mathbf{R}^n$  and a point  $\bar{x} \in Q$ . We let the *local dimension* of  $Q$  at  $\bar{x}$  be

$$\dim_Q(\bar{x}) := \inf_{\epsilon > 0} \dim(Q \cap B_\epsilon(\bar{x})).$$

It is not difficult to see that there exists a real number  $\bar{\epsilon} > 0$  such that for every real number  $0 < \epsilon < \bar{\epsilon}$ , we have  $\dim_Q(\bar{x}) = \dim(Q \cap B_\epsilon(\bar{x}))$ .

There is a straightforward connection between local dimension and dimension of strata. This is the content of the following proposition.

**Proposition 5.6.** [10, Proposition 3.4] *Consider a semi-algebraic set  $Q \subset \mathbf{R}^n$  and a point  $\bar{x} \in Q$ . Let  $\{M_i\}$  be any stratification of  $Q$ . Then we have the identity*

$$\dim_Q(\bar{x}) = \max_i \{\dim M_i : \bar{x} \in \text{cl } M_i\}.$$

**Definition 5.7** (Maximal strata). Given a stratification  $\{M_i\}$  of a semi-algebraic set  $Q \subset \mathbf{R}^n$ , a stratum  $M$  is *maximal* if it is not contained in the closure of any other stratum.

Using the defining property of a stratification, we can equivalently say that given a stratification  $\{M_i\}$  of a semi-algebraic set  $Q \subset \mathbf{R}^n$ , a stratum  $M_i$  is maximal if and only if it is disjoint from the closure of any other stratum.

**Proposition 5.8.** [10, Proposition 3.7] Consider a stratification  $\{M_i\}$  of a semi-algebraic set  $Q \subset \mathbf{R}^n$ . Then given any point  $\bar{x} \in Q$ , there exists a maximal stratum  $M$  satisfying  $\bar{x} \in \text{cl } M$  and  $\dim M = \dim_Q(\bar{x})$ .

Semi-algebraic methods have recently found great uses in set-valued analysis. See for example [9, 11, 18, 16, 17]. A fact that will be particularly useful for us is that semi-algebraic set-valued mappings are “generically” inner-semicontinuous.

**Proposition 5.9.** [11, Proposition 2.28, 2.30] Consider a semi-algebraic, set-valued mapping  $G: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ . Then there exists a stratification of  $\text{dom } G$  into finitely many semi-algebraic manifolds  $\{M_i\}$  such that on each stratum  $M_i$ , the mapping  $G$  is inner-semicontinuous and the dimension of the images  $F(x)$  is constant. If in addition  $F$  is closed-valued, then we can ensure that the restriction  $G|_{M_i}$  is also outer-semicontinuous for each index  $i$ .

For a more refined result along the lines of Proposition 5.9, see [9, Theorem 28]. The following result is standard.

**Proposition 5.10.** [7, Theorem 3.18] Consider a semi-algebraic, set-valued mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ . Suppose there exists an integer  $k$  such that  $F(x)$  is  $k$ -dimensional for each point  $x \in \text{dom } F$ . Then the equality,

$$\dim \text{gph } F = \dim \text{dom } F + k,$$

holds.

We will need a version of Proposition 5.10 that pertains to local dimension.

**Proposition 5.11.** Consider a semialgebraic mapping  $F: \mathbf{R}^n \rightrightarrows \mathbf{R}^m$  that is inner-semicontinuous on its domain. Suppose that there exist constants  $k$  and  $l$  such that for each pair  $(x, v) \in \text{gph } F$ , we have

$$\dim_{\text{dom } F} x = k, \quad \dim_{F(x)} v = l.$$

Then  $\text{gph } F$  has local dimension  $k + l$  around every pair  $(x, v) \in \text{gph } F$ .

*Proof.* Let  $\pi: \mathbf{R}^n \times \mathbf{R}^m \rightarrow \mathbf{R}^n$  be the canonical projection onto  $\mathbf{R}^n$ . Consider any stratification  $\mathcal{A}$  of  $\text{gph } F$  and let  $M \in \mathcal{A}$  be any maximal stratum. Clearly

$$\dim M \leq \dim \text{gph } F \leq k + l.$$

Consider an arbitrary point  $x \in \pi(M)$ . Since  $M$  is maximal, the set  $M \cap (\{x\} \times \mathbf{R}^m)$  is open relative to  $\text{gph } F \cap (\{x\} \times \mathbf{R}^m)$ . Furthermore, since  $\dim_{F(x)} v = l$  for each vector  $v \in F(x)$ , it easily follows that  $\dim M \cap (\{x\} \times \mathbf{R}^m) = l$ .

We now claim that  $\dim \pi(M) = k$ . Indeed suppose this is not the case, that is the strict inequality  $\dim \pi(M) < k$  holds. Since  $\dim_{\text{dom } F} x = k$ , we deduce that there exists a sequence  $x_i \rightarrow \bar{x}$ , with  $x_i \in \text{dom } F$  and  $x_i \notin \pi(M)$  for each index  $i$ . Since  $F$  is inner-semicontinuous on  $M$ , we deduce

$$M \cap (\{x\} \times \mathbf{R}^m) \subset \limsup_{i \rightarrow \infty} \{x_i\} \times F(x_i),$$

which contradicts maximality of  $M$ . Consequently, using Proposition 5.10, we deduce  $\dim M = k + l$ . Since  $M$  was an arbitrary maximal stratum, the result follows by an application of Proposition 5.8.  $\square$

## 5.2 Local dimension of semi-algebraic subdifferential graphs.

We begin with a definition.

**Definition 5.12** (Subjects). For a function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , the *limiting subject* is given by

$$[\partial f] := \{(x, f(x), v) : v \in \partial f(x)\}.$$

Subjects corresponding to the other subdifferentials are defined analogously.

Much like  $f$ -attentive convergence, subjects are useful for keeping track of variational information in absence of continuity. In this section, we build on the following theorem. This result and its consequences for generic semi-algebraic optimization problems are discussed extensively in [11].

**Theorem 5.13.** [11, Theorem 3.6] *Let  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  be a proper semi-algebraic function. Then the subjects  $[\partial_P f]$ ,  $[\hat{\partial} f]$ ,  $[\partial f]$  and  $[\partial_c f]$  have dimension exactly  $n$ .*

An immediate question arises: Can the four subjects have local dimension smaller than  $n$  at some of their points? In a recent paper [10], the authors showed that this indeed may easily happen for  $[\partial_c f]$ . Remarkably the authors showed that the subjects  $[\partial_P f]$ ,  $[\hat{\partial} f]$ , and  $[\partial f]$  of a lower-semicontinuous, semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$  do have uniform local dimension  $n$ . The significance of this result and the relation to Minty's theorem were also discussed. In this section, we provide a much simplified proof of this rather striking fact (Theorem 5.16). The main tool we use is the following accessibility lemma, which is a special case of Lemma 3.5. Since the proof is much simpler than that of Lemma 3.5, we include the full argument below.

**Lemma 5.14** (Accessibility). *Consider a closed set  $Q \subset \mathbf{R}^n$ , a manifold  $M \subset Q$ , and a point  $\bar{x} \in M$ . Recall that the inclusion  $N_Q^P(\bar{x}) \subset N_M(\bar{x})$  holds. Suppose that a proximal normal vector  $\bar{v} \in N_Q^P(\bar{x})$  lies in the boundary of  $N_Q^P(\bar{x})$ , relative to the linear space  $N_M(\bar{x})$ . Then there exist sequences  $x_i \rightarrow \bar{x}$  and  $v_i \rightarrow \bar{v}$ , with  $v_i \in N_Q^P(x_i)$ , and so that all the points  $x_i$  lie outside of  $M$ .*

*Proof.* There exists a real number  $\lambda > 0$  so that  $\bar{x} + \lambda \bar{v}$  lies in the prox-normal neighborhood  $W$  of  $M$  at  $\bar{x}$  and such that the equality  $P_Q(\bar{x} + \lambda \bar{v}) = \bar{x}$  holds. Consider any sequence  $v_i \in \mathbf{R}^n$  satisfying

$$v_i \rightarrow \bar{v}, v_i \in N_M(\bar{x}), v_i \notin N_Q^P(\bar{x}).$$

Choose arbitrary points  $x_i \in P_Q(\bar{x} + \lambda v_i)$ . We have

$$(\bar{x} - x_i) + \lambda v_i \in N_Q^P(x_i).$$

We deduce  $x_i \neq \bar{x}$ . Clearly, the sequence  $x_i$  converges to  $\bar{x}$ . We claim  $x_i \notin M$  for all sufficiently large indices  $i$ . Indeed, if it were otherwise, then for large  $i$ , the points  $\bar{x} + \lambda v_i$  would lie in  $W$  and we would have  $x_i \in P_M(\bar{x} + \lambda v_i) = \bar{x}$ , which is a contradiction. Thus we have obtained a sequence  $(x_i, \frac{1}{\lambda}(\bar{x} - x_i) + v_i) \in \text{gph } N_Q^P$ , with  $x_i \notin M$ , and satisfying  $(x_i, \frac{1}{\lambda}(\bar{x} - x_i) + v_i) \rightarrow (\bar{x}, \bar{v})$ .  $\square$

The following is now immediate.

**Corollary 5.15.** *Consider a lower semicontinuous function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ , a manifold  $M \subset \mathbf{R}^n$ , and a point  $\bar{x} \in M$ . Suppose that  $f$  is smooth on  $M$  and the strict inequality  $\dim \partial_P f(\bar{x}) < \dim N_M(\bar{x})$  holds. Then for every vector  $\bar{v} \in \partial_P f(\bar{x})$ , there exist sequences  $(x_i, f(x_i), v_i) \rightarrow (\bar{x}, f(\bar{x}), \bar{v})$ , with  $v_i \in \partial_P f(x_i)$ , and so that all the points  $x_i$  lie outside of  $M$ .*

*Proof.* From the strict inequality, one can easily see that the normal cone  $N_{\text{epi } f}^P(\bar{x}, f(\bar{x}))$  has empty interior relative to the normal space  $N_{\text{gph}}(\bar{x}, f(\bar{x}))$ . An application of Lemma 5.14 completes the proof.  $\square$

We can now prove the main result of this section.

**Theorem 5.16.** *Consider a lower-semicontinuous, semi-algebraic function  $f: \mathbf{R}^n \rightarrow \overline{\mathbf{R}}$ . Then the subjects  $[\partial_P f]$ ,  $[\hat{\partial} f]$ , and  $[\partial f]$  have constant local dimension  $n$  around each of their points.*

*Proof.* We first prove the theorem for the subset  $[\partial_P f]$ . Consider the semi-algebraic set-valued mapping

$$F(x) := \{f(x)\} \times \partial_P f(x),$$

whose graph is precisely  $[\partial_P f]$ . By Propositions 5.3 and 5.9, we may stratify the domain of  $F$  into finitely many semi-algebraic manifolds  $\{M_i\}$ , so that on each stratum  $M_i$ , the mapping  $F$  is inner-semicontinuous, the images  $F(x)$  have constant dimension, and  $f$  is smooth. Consider a triple  $(x, f(x), v) \in [\partial_P f]$ . We prove the theorem by induction on the dimension of the strata  $M$  in which the point  $x$  lies. Clearly the result holds for the strata of dimension  $n$ , if there are any. As an inductive hypothesis, assume that the theorem holds for all points  $(x, f(x), v) \in [\partial_P f]$  with  $x$  lying in strata of dimension at least  $k$ , for some integer  $k \geq 1$ .

Now consider a stratum  $M$  of dimension  $k - 1$  and a point  $x \in M$ . If  $\dim F(x) = n - \dim M$ , then recalling that  $F$  is inner-semicontinuous on  $M$  and applying Proposition 5.11, we see that the set  $\text{gph } F|_M$  has local dimension  $n$  around  $(x, f(x), v)$  for any  $v \in \partial_P f(x)$ . The result follows in this case.

Now suppose  $\dim F(x) < n - \dim M$ . Then, by Corollary 5.15, for such a vector  $v$ , there exists a sequence  $(x_i, f(x_i), v_i) \rightarrow (x, f(x), v)$  satisfying  $(x_i, f(x_i), v_i) \in [\partial_P f]$  and  $x_i \notin M$  for each index  $i$ . Restricting to a subsequence, we may assume that all the points  $x_i$  lie in a stratum  $K$  satisfying  $\dim K \geq k$ . By the inductive hypothesis, we deduce

$$\dim_{[\partial_P f]}(x, f(x), v) \geq \limsup_{i \rightarrow \infty} \dim_{[\partial_P f]}(x_i, f(x_i), v_i) = n.$$

This completes the proof of the inductive step and of the theorem for the subset  $[\partial_P f]$ .

Now observe that  $[\partial_P f]$  is dense in  $[\hat{\partial} f]$  and in  $[\partial f]$ . It follows that  $[\hat{\partial} f]$  and  $[\partial f]$  also have local dimension  $n$  around each of their points.  $\square$

Surprisingly Theorem 5.16 may fail in the Clarke case, even for Lipschitz continuous functions.

**Example 5.17.** Consider the function  $f: \mathbf{R}^3 \rightarrow \mathbf{R}$ , defined by

$$f(x, y, z) = \begin{cases} \min\{x, y, z^2\} & \text{if } (x, y, z) \in \mathbf{R}_+^3 \\ \min\{-x, -y, z^2\} & \text{if } (x, y, z) \in \mathbf{R}_-^3 \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\bar{x} \in \mathbf{R}^n$  be the origin and let  $\Gamma := \text{conv}\{(1, 0, 0), (0, 1, 0), (0, 0, 0)\}$ . One can check that the local dimension of  $\text{gph } \partial_c f$  at  $(\bar{x}, \bar{v})$  is two for any vector  $\bar{v} \in (\text{conv}(\Gamma \cup -\Gamma)) \setminus (\Gamma \cup -\Gamma)$ . For more details see [10, Example 3.11].

## References

- [1] J.M. Borwein, J.V. Burke, and A.S. Lewis. Differentiability of cone-monotone functions on separable Banach space. *Proc. Amer. Math. Soc.*, 132(4):1067–1076 (electronic), 2004.
- [2] J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis*. Springer Verlag, New York, 2005.
- [3] J.V. Burke, A.S. Lewis, and M.L. Overton. Approximating subdifferentials by random sampling of gradients. *Mathematics of Operations Research*, 27(3):567–584, 2002.
- [4] J.V. Burke, A.S. Lewis, and M.L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM J. on Optimization*, 15:751–779, March 2005.
- [5] F.H. Clarke. Generalized gradients and applications. *Transactions of the American Mathematical Society*, 205:247–262, Apr. 1975.
- [6] F.H. Clarke, Yu. Ledyev, R.I. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory*. Texts in Math. 178, Springer, New York, 1998.
- [7] M. Coste. *An Introduction to  $\phi$ -minimal Geometry*. RAAG Notes, 81 pages, Institut de Recherche Mathématiques de Rennes, November 1999.
- [8] M. Coste. *An Introduction to Semialgebraic Geometry*. RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.

- [9] A. Daniilidis and C.H.J. Pang. Continuity and differentiability of set-valued maps revisited in the light of tame geometry. *J. Lond. Math. Soc. (2)*, 83(3):637–658, 2011.
- [10] D. Drusvyatskiy, A. D. Ioffe, and A.S. Lewis. The dimension of semialgebraic subdifferential graphs. *Nonlinear Anal.*, 75(3):1231–1245, 2012.
- [11] D. Drusvyatskiy and A.S. Lewis. Semi-algebraic functions have small subdifferentials. *To appear in Mathematical Programming, Ser. B*, 2012.
- [12] A. Marigonda G. Colombo. Singularities for a class of non-convex sets and functions, and viscosity solutions of some Hamilton-Jacobi equations. *Journal of Convex Analysis*, 15(1):105 – 129, 2008.
- [13] P. Wolenski G. Colombo, A. Marigonda. The clarke generalized gradient for functions whose epigraph has positive reach. *To appear in Mathematics of Operations Research*, 2012.
- [14] W.L. Hare and A.S. Lewis. Identifying active constraints via partial smoothness and prox-regularity. *J. Convex Anal.*, 11(2):251–266, 2004.
- [15] W.L. Hare and A.S. Lewis. Estimating tangent and normal cones without calculus. *Math. Oper. Res.*, 30(4):785–799, 2005.
- [16] A.D. Ioffe. A Sard theorem for tame set-valued mappings. *J. Math. Anal. Appl.*, 335(2):882–901, 2007.
- [17] A.D. Ioffe. Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems. *Proc. Amer. Math. Soc.*, 136(9):3111–3119, 2008.
- [18] A.D. Ioffe. An invitation to tame optimization. *SIAM Journal on Optimization*, 19(4):1894–1917, 2009.
- [19] J.M. Lee. *Introduction to Smooth Manifolds*. Springer, New York, 2003.
- [20] G.J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29:341–346, 1962.
- [21] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory*. Grundlehren der mathematischen Wissenschaften, Vol 330, Springer, Berlin, 2006.
- [22] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation II: Applications*. Grundlehren der mathematischen Wissenschaften, Vol 331, Springer, Berlin, 2006.
- [23] R.A. Poliquin and R.T. Rockafellar. Amenable functions in optimization. In *Nonsmooth optimization: methods and applications (Erice, 1991)*, pages 338–353. Gordon and Breach, Montreux, 1992.
- [24] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Trans. Amer. Math. Soc.*, 348:1805–1838, 1996.
- [25] S.M. Robinson. A point-of-attraction result for Newton’s method with point-based approximations. *Optimization*, 60(1-2):89–99, 2011.
- [26] S.M. Robinson. Equations on monotone graphs. *Mathematical Programming*, pages 1–53, 2012.
- [27] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [28] R.T. Rockafellar. Clarke’s tangent cones and the boundaries of closed sets in  $\mathbf{R}^n$ . *Nonlinear Analysis: Theory, Methods, and Applications*, 3(1):145 – 154, 1979.
- [29] R.T. Rockafellar and R.J-B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.
- [30] L. van den Dries and C. Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84:497–540, 1996.